DEFENCE
SCIENCE & TECHNOLOGY

Suitability of Correlation Arrays and Superresolution for Minehunting Sonar

David G. Blair

DSTO-TN-0443

DEFENCE
SCIENCE & TECHNOLOGY

Suitability of Correlation Arrays and Superresolution for Minehunting Sonar

David G. Blair

Maritime Operations Division
Systems Sciences Laboratory

DSTO-TN-0443

ABSTRACT

Towards the possible development of new sonars, particularly a hull-mounted minehunting sonar, work has been done in two areas. The first area concerns arrays—to be called 'correlation arrays'—that are based on the principle used by correlation telescopes in radio astronomy. The application of this principle might enable a large reduction in the number of sonar array elements. In this area, key results from the literature are reported and some concepts are elucidated with the aid of mathematics. The second area is that of superresolution (SR) and superdirectivity. The former consists of methods by which resolution is obtained that is better than that given by the Rayleigh criterion. In this second area, a more complete literature survey is carried out. A number of SR techniques are briefly described, including Capon's method and an autoregressive moving average (ARMA) model. An important problem for the application of the above techniques to sonar is that both SR and correlation arrays apparently require a high signal-to-noise ratio. For both SR and correlation arrays, the application to sonar faces a further key problem. The problem is that correlations usually exist between the return signals from different targets, and these correlations lead to artefacts and other image defects. Known techniques for 'decorrelating' targets are described.

RELEASE LIMITATION

Approved for public release

Published by

DSTO Systems Sciences Laboratory
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-012-410
July 2002

APPROVED FOR PUBLIC RELEASE

Suitability of Correlation Arrays and Superresolution for Minehunting Sonar

Executive Summary

As part of the Navy's minehunting program, the need has been identified to acquire, if possible, improved hull-mounted sonars. These might, in particular, survey an area faster, and/or enable the detection and classification of suspected mines to occur at a greater range. Towards developing such a sonar, a project called 'Minefinder' has been set up. As part of Minefinder, work, reported here, has been carried out in two areas.

The first area concerns arrays—to be called 'correlation arrays'—that are based on the principle used by correlation telescopes in radio astronomy. The best-known of these devices is the cross-shaped array, constructed from two arms and commonly called the Mills cross. In some sense, such an array of $M+N$ elements performs as well as a conventional array of MN elements. Thus the application of the principle might enable a large reduction in the number of sonar array elements. In this area, key results from the literature are reported and some concepts are elucidated with the aid of mathematics. Key problems are identified: in particular, two problems described in the last paragraph of this summary.

The second area is that of superresolution (SR), consisting of methods by which resolution is obtained that is better than that given by the Rayleigh criterion. (The latter states that angular resolution is approximately equal to (wavelength)/(dimension of array).) Potentially SR would produce better sonar images at any given range. In this second area, a more complete literature survey is carried out. A number of SR techniques are briefly described, including Capon's method and an autoregressive moving average (ARMA) model. In SR, while each method has had successes, general conclusions as to the goodness of the methods are in general lacking in the literature. A book by Steinberg and Subbaram is of particular interest, since it, like the Minefinder sonar, is concerned with active imaging.

An important problem for the application of the above techniques to sonar is that both SR and correlation arrays apparently require a high signal-to-noise ratio. For both SR and correlation arrays, the application to sonar faces a further key problem. The problem is that correlations usually exist between the return signals from different targets; that is, these signals bear a certain resemblance to each other. These correlations cause the image quality to become degraded. Success therefore requires some method for 'decorrelating' the targets. Methods that achieve this under certain circumstances are described.

Contents

1. Introduction

This work was carried out as part of a project, called 'Minefinder,' within the Maritime Operations Division, DSTO. The aim of this project, in the first instance, is to do the preliminary design for a hull-mounted sonar to detect and classify sea mines. That design is to be followed by experimentation to prove any new concepts and to separate feasible from non-feasible designs. This could lead to a state-of-the-art sonar being constructed by a commercial firm.

It was decided to look at some new options for the sonar that might improve performance, including options that have generally been avoided in the minehunting context. One possibility is to beamform in elevation as well as in azimuth. As well as giving information on the elevation of targets, this vertical resolution might help avoid problems with multipaths.

It is intended to devote effort also to the use of coded signals, especially multiple, orthogonal (or almost orthogonal) pseudo-random sequences. By allowing many signals to be in the water at once, coded signals could help to remove the limitation placed on the rate of data acquisition by the long time taken for the signal to travel to the furthest range and return. The coded beams could be used to produce several narrow beams with different elevation angles. But they could also be used to send a few signals to probe at short ranges at the same time as a differently coded signal probes at longer ranges.

The present report is in large part a literature survey, but also contains some analysis of the implications of the survey for the proposed sonar.

The first broad area covered by the report is inspired by work in radio astronomy. First, in long-baseline interferometry, two widely separated arrays are used to obtain as good a resolution as obtained from a single array joining the two. This concept might be used to obtain high resolution by mounting two arrays out to the side in 'free' water, one on each side of the ship's bow. Second, consider the cross-shaped array, constructed from two subarrays in the form of two arms perpendicular to each other, one subarray having M elements and the other having N. When this arrangement is used as a correlation telescope, in some sense the $M+N$ elements achieve as good a result as the MN elements that a conventional array would use. Applied to sonar, this concept might greatly reduce the number of elements needed.

The second broad area, superresolution, is inspired by a general problem posed by bow-mounted sonars. From the viewpoint of sonar performance, it is desirable for the array to be long in the athwartships direction. But a long array produces high drag, which is quite undesirable. As a result such sonars are generally made less than 2 m long. It is known that under certain conditions the resolution achievable with an array can be improved beyond that given by the Rayleigh criterion; this phenomenon is called superresolution (SR). This suggests the question: Can superresolution techniques be used in our application to increase the azimuthal resolution for an array of fixed length? The aim in this second broad area is to draw together information from the literature with a view to answering this question. It should also be noted that,

even if the array length *can* be made considerably larger than 2 m despite the drag problem, SR would still be of use, since better resolution is always desirable.

It turns out that the two strands of the report are closely connected. This is because the success of superresolution techniques usually depends on there being no correlation between the returns from different targets. (Here the 'correlation' is the result obtained by multiplying the two returns together and averaging over time.) Similarly, with the application of radio-astronomy techniques to sonar (correlation arrays), we have exactly the same precondition for success, namely, that there is no correlation between the returns from different targets (or sources in the passive case). The targets in our problem are in fact correlated; hence a very relevant part of the literature deals with means for 'decorrelating' the targets.

Sections 2 to 4 deal with correlation arrays. Sections 5 to 12 deal with superresolution; however, Section 11, on methods of decorrelation, is thought to have application also to correlation arrays. Conclusions are given in Section 13.

2. Introduction to Correlation Arrays

2.1 General

This part of the work—correlation arrays—began with a consideration of the cross-shaped array, constructed from two subarrays in the form of two arms perpendicular to each other— often referred to loosely as a 'Mills cross.' It was noted that the use of this arrangement is often beneficial, both in radio astronomy and in sonar.

In view of the common confusion, clarification is needed regarding how such a device works. The original Mills cross [Christiansen and Hogbom 1969, pp. 149–154], used in radio astronomy, involves forming the correlation of pairs of signals (one from each arm). That device belongs to a class called *correlation telescopes*; these are discussed in Section 3. The term *correlation array* will be used for sonar devices (and other terrestrial devices) that are based on the same principle. Note that such arrays can have other geometries besides the cross shape.

There are two other ways in which the cross-shaped geometry might be used—one successful and one not. These are described in Section 2.2.

Section 4 discusses the challenges posed by the application of the correlation array principle to sonar. In particular, the problem of *correlated target strengths* is identified. This problem will again come to the fore when superresolution is discussed. In Section 11, possible solutions to the 'decorrelation' problem in the context of superresolution are discussed. It would appear that at least some of these methods can be adapted to the case of correlation arrays. Conclusions on both correlation arrays and superresolution are gathered together in Section 13.

The literature on correlation arrays is almost totally on systems that are both narrowband and passive. Hence, in this report's discussion of correlation arrays (Sections 2 to 4), the major part is on narrowband, passive systems. However Sections

4.2 to 4.4 present 'slightly new' work that is about active systems and is 'wideband' in a useful sense.

2.2 Possible Ways of Using the Cross-Shaped Array

Given a set of elements that forms a cross shape, there are at least three ways in which those elements could be used. The first way is the correlation telescope mentioned above.[1] A second way is to use the elements as a normal sonar receiving array. However, the two-dimensional (2-D) array thus defined is quite unsatisfactory, for the following reason. Since the grading, or weighting, function is the sum of the two grading functions, the far-field beam pattern is the sum of the Fourier transforms of the two grading functions. For simplicity, consider a system in which each arm is an aperture (continuous array), so that the overall grading function [Steinberg 1976] is

$$g(x,y) = \mathrm{rect}(x/L)\delta(y) + \delta(x)\mathrm{rect}(y/L) \tag{2.1}$$

where L is the length of each of the two arms. Then the (amplitude) beam pattern is

$$G(l,m) = L\mathrm{sinc}\left(L\lambda^{-1}l\right) + L\mathrm{sinc}\left(L\lambda^{-1}m\right) \tag{2.2}$$

where the direction cosines l and m are given in terms of the coordinates (x,y,z) of the image point by $l = x/r$, $m = y/r$, where $r = \sqrt{x^2+y^2+z^2}$ is the range. (An image point is a point at which an image intensity is to be calculated.) (The expressions (2.1) and (2.2) have not been normalised.) Then along the x axis, when we move from the peak of the pattern to a point *many* sidelobes out along the x axis, the image intensity drops off by only 6 dB. Such a pattern is clearly unsatisfactory.

A third way of applying the cross-shape arrangement is to use one of the arms on transmit and the other arm on receive. This method works, because the combined beam pattern is then the *product* of the two component beam patterns. The latter feature results in the combined beam pattern intensity being small unless the transverse displacement of the image point (from the point target) is small in *both* the x and the y directions. Here 'small' means 'not significantly greater than the beamwidth.'

Thus, of the three methods described, the first and the third are sound while the second is unsound. Yet the belief is widely held that the second method has been used, commonly and successfully, in sonar. This cannot be. As far as the author is aware, the cross-shaped arrays that have been used in sonar all utilise the third method.

---

[1] As described in Section 3, the correlation telescope itself can be considered to come in a 'traditional' and an 'updated' version.

3. Correlation Telescopes in Radio Astronomy

3.1 General

In radio astronomy, instruments based on the cross-shaped array have been highly successful, even though used on receive only. This is possible because, in radio astronomy, the cross-shaped array is always used as a *correlation telescope*—the 'first way' above. (Details are given in Christiansen and Hogbom [1969] (to be called C&H), (pp. 103–170); useful background is given in pp. 1–42. Techniques of radio astronomy are also discussed in Perley *et al.* [1989].) The cross-shaped array used in this fashion is indeed the original 'Mills cross' (C&H, pp. 149–154), also called in radio astronomy the *cross antenna*. The cross shape is not unique: there are several geometries that have been used as the basis for a correlation telescope. These other telescopes differ from the cross antenna in a few ways; a brief discussion of them is given later in this Section 3.

C&H (pp. 14, 104, 115, 169) emphasise that, for a correlation telescope to count as *filled*, all that is needed is that a sufficient set of *vector spacings of pairs of elements* be used. The 'sufficiency' condition is that the $\lambda/2$ criterion applies to these spacings, just as if the vector spacings themselves formed an ordinary 2-D array. In other words, if all those vector spacings were plotted as vectors from a common origin, the array of end-points so formed must be no more sparse than a square array of spacing $\lambda/2$. *From this point of view, a filled square array with spacing $\lambda/2$—with correlation performed on all pairs—would have a lot of redundancy.* The cross antenna makes use of this redundancy by removing most of the elements. The cross antenna—or, to be precise, an even more sparse arrangement in the shape of a 'T'—has no redundancy. It must be stressed that the assertion regarding 'a sufficient set of vector spacings' holds only for *correlation* telescopes; the latter qualification is not always clear from the early part of the book of C&H.

In the remainder of Section 3, the cross antenna will be discussed in some detail. The main discussion (Section 3.2) concerns what will be called the 'updated' version. A typical version of the cross antenna up to the time of the book by C&H [1969] will be called the 'traditional' version. The difference lies in the fact that, since 1969, it has become possible to use software to perform much of the work that was previously carried out by analogue hardware devices. Thus greater flexibility can now be achieved. The 'updated' version of the cross antenna, based on such a replacement, exhibits a strong similarity to modern sonar devices. Having discussed the updated version, we shall then note (Section 3.3) that the traditional version of the cross antenna is indeed basically the same instrument, being based on the same mathematics.

3.2 Updated Cross Antenna; Telescope Transfer Function

The 'updated' cross antenna would work as follows. The antenna or array is made up of two subarrays, each being an arm of the array. Let n label a typical element in subarray 1 and p label a typical element in subarray 2. The array is sensitive to

incoming radiation over a narrow band $\Delta\nu$ about a central frequency ν. The voltage output from element n in subarray 1 is E_n, where the latter, $E_n = E_n(t)$, actually denotes the complex envelope of the analytic signal. (The envelope is formed by removing the factor $e^{j2\pi\nu t}$.) E_n varies with time, but the variation is slow on the scale of $1/\nu$. Similarly the voltage output from element p in subarray 2 is F_p. At each time, the product $E_n F_p^*$ is formed (where the star denotes complex conjugate). For each pair (n, p), this product is averaged over time; thus the *correlation* of E_n with F_p is obtained.

Now consider beamforming. In correlation telescopes, such as the cross antenna, the fundamental unit is not a single element but a pair of elements (n, p), one from each of the two subarrays. The steering factor to be applied is no longer

$$\exp\{j2\pi\lambda^{-1}[lx_n + my_n]\}$$

but (it being remembered that the product $E_n F_p^*$ contains a complex conjugate)

$$\exp\{j2\pi\lambda^{-1}[l(x_n - x_p) + m(y_n - y_p)]\} \tag{3.1}$$

Here, for example, (x_n, y_n) are the coordinates of the nth element.[2] l and m are the direction cosines of the steering direction $\hat{r}$; specifically $l = \cos(\hat{r}, \hat{x})$, $m = \cos(\hat{r}, \hat{y})$, where $(\mathbf{A}, \mathbf{B})$ denotes the angle between the vectors $\mathbf{A}$ and $\mathbf{B}$, and $\hat{x}$ is a unit vector along the x axis. (In (3.1), each x_p and each y_n is actually zero for the cross antenna.) Thus in beamforming, the position vector as a basic concept is replaced by the *vector spacing*, $\mathbf{r}_n - \mathbf{r}_p$, from the one element to the other. The vector language is appropriate here, since the square bracket in (3.1) can be written as the scalar product

$$(l, m) \cdot (\mathbf{r}_n - \mathbf{r}_p) = (l, m) \cdot \mathbf{u}_{np} \tag{3.2}$$

Here $\mathbf{u}_{np} = \mathbf{r}_n - \mathbf{r}_p$ is the spacing associated with the pair (n, p).

We can build up an analogy between the cross antenna (or more generally the correlation telescope) and the normal array. We have just seen that in the analogy, the spacing $\mathbf{u}_{np}$ replaces the position $\mathbf{r}_n$. As we shall see, the grading function $g(x, y)$ is replaced by the *telescope transfer function* (or correlation function), particularly in regard to their connection with the beam pattern. The telescope transfer function is defined by

---

[2] Whether j should be replaced by $-j$ in equations such as (3.1), rests on a convention. In the present discussion the convention of C&H has been followed. If the opposite convention is adopted, the observable results are the same.

$$c(u,v) = \text{const. } g_1(x,y) \otimes g_2^*(x,y) \tag{3.3}$$

where g_1 and g_2 are the grading functions of the respective subarrays and $\otimes$ denotes the correlation operation defined by

$$f(x,y) \otimes g(x,y) = \iint f(x,y) g(x-u, y-v) dxdy \tag{3.4}$$

Because of the form of (3.4), $c(u,v)$ may be viewed as the total complex weight associated with the vector spacing (u,v), i.e. the superposition of all products fg consistent with that vector spacing. Then, as shown by C&H [1969, p. 116], for a correlation telescope, the intensity beam pattern—or intensity directivity function of the array when steered to the zenith—is

$$A(l,m) = \iint c(u,v) \exp\left[j2\pi\lambda^{-1}(lu+mv)\right] dudv \tag{3.5}$$

Thus *the beam pattern* $A(l,m)$ *and the transfer function* $c(u,v)$ *form a Fourier transform pair.* When we compare this result with the normal array, in which the amplitude beam pattern $B(l,m)$ and the grading function form a Fourier transform pair:

$$B(l,m) = \text{const.} \iint g(x,y) \exp\left[j2\pi\lambda^{-1}(lx+my)\right] dxdy \tag{3.6}$$

[Steinberg 1976], the analogy between $g(x,y)$ and $c(u,v)$ is demonstrated. As we shall see (assumptions 2 and 3 in Section 4.1 below), radio astronomers are fortunate in that the relationship (3.5) is essentially exact.

Note that, in the overall analogy that we have been developing, the *amplitude* beam pattern of the normal array is replaced by the *intensity* beam pattern in the radio astronomy case. Note also that, in radio astronomy, quite generally the intensity of the image equals the true intensity pattern $I_s(l,m)$ of the sky convolved with the beam pattern (3.5). Thus the image intensity is

$$I(l,m) = \iint I_s(l',m') A(l-l', m-m') dl'dm' \tag{3.7}$$

or, in the form that assumes discrete sources,

$$I(l,m) = \sum_j \left|a_j\right|^2 A(l-l_j, m-m_j) \tag{3.8}$$

where $\left|a_j\right|^2$ is proportional to the intensity received at the antenna from the j th source and the latter is in the direction (l_j, m_j).

Equation (3.5) may also be written as

$$A(l,m) = \Gamma(l/\lambda, m/\lambda) \tag{3.9}$$

where Γ is the inverse Fourier transform of the transfer function $c(u,v)$ (see definition of Fourier transform at Eqn (4.2) below).

For the cross antenna, characterised by a continuous, uniformly graded antenna with equal arms, the evaluation of Equation (3.5) yields[3] for the beam pattern

$$A(l,m) = \text{const.}\left|\operatorname{sinc}(\lambda^{-1}Ll)\operatorname{sinc}(\lambda^{-1}Lm)\right| \tag{3.10}$$

[3] for the *cosine* component (refer to C&H).

(C&H, p. 149). Note that a *product* of the component beam patterns is now obtained; contrast the *sum* in (2.2). For this reason, outside the beamwidth, the sidelobes fall off rapidly in *all* directions and the instrument is very satisfactory. Note also that, although the left side of (3.10) is an intensity, the sinc functions on the right side are *not* squared.

3.3 Traditional Cross Antenna

We now discuss the *traditional* cross antenna in radio astronomy. Here, while each subarray consists of dipoles—the analogue of elements in sonar—the voltages from these dipoles are not individually recorded, nor are any correlations carried out between an individual dipole voltage and another dipole voltage. Instead, for the first arm, after the analogue implementation of phase delays to produce steering in one of the two angular coordinates, the total voltage on the arm is obtained and treated as a single unit. Similarly for the other arm, which is steered in the other angular coordinate. The correlation of these two totals is then obtained by analogue means.

The traditional telescope's instantaneous output is thus given by the first line of the following:

$$\begin{aligned} V &= \left\{\sum_n E_n \exp\left[j2\pi\lambda^{-1}(l,m)\cdot \mathbf{r}_n\right]\right\}\left\{\sum_p F_p \exp\left[j2\pi\lambda^{-1}(l,m)\cdot \mathbf{r}_p\right]\right\}^* \\ &= \sum_n \sum_p E_n F_p^* \exp\left[j2\pi\lambda^{-1}(l,m)\cdot \mathbf{u}_{np}\right] \end{aligned} \tag{3.11}$$

where $\mathbf{u}_{np}$ is defined below (3.2). (Actually the right-hand side of Eqn (3.11) is to be averaged over time to eliminate uncorrelated contributions to the instantaneous product.) Removal of the curly brackets gives the second line. The second line shows that the resulting output from the traditional telescope is the same as if the following two steps were carried out: (i) the voltages are multiplied *within each element pair* to yield $E_n F_p^*$; and (ii) delay-and-add beamforming is then applied to the product by performing a phase shift based on the vector spacing $\mathbf{u}_{np}$ and then adding. But this is the very beamforming process described for the *updated* antenna at Equation (3.1). Thus *in essence the traditional telescope behaves in the same way as the updated telescope.*

Of course, an important difference in practice is that, in a given correlation run as described in the first paragraph, the traditional telescope yields an image intensity for just one direction in the sky (one resolution cell). By contrast, the updated telescope would record all the voltage streams from the *one* correlation run and, by repeated reprocessing of the same data, would yield the image intensity for *all* relevant directions.

3.4 Compound Grating Antenna

Of the other correlation telescopes discussed by C&H [1969], the most interesting for present purposes is the *compound grating antenna* or compound interferometer (pp. 132–144). Only a brief description will be given. First, normally the elements are arranged, not in a cross, but all in the same row. Second, each element is now, not a dipole, but a paraboloidal dish (or similar device), and therefore has very considerable directionality of its own. The latter directionality is enhanced (in one direction) by the array arrangement. All this forms the first 'subarray.' The second 'subarray' may be a *single dish* placed in line with the first subarray. However, in the most interesting case, the second subarray is *another row of elements* in line with the first.

The above discussion of the updated and traditional telescopes still applies in essence to the compound grating antenna. But there are two complications, as follows. (i) Since each element now has directionality, an extra factor for this must be inserted in the beam pattern. (ii) Steering of the array is not accomplished by introducing, electrically or computationally, the large phase delays implied by Equation (3.11). Rather, first, each dish is *mechanically* steerable; it may be said that most of the steering is accomplished by steering each element mechanically. Second, after the first step, it remains to adjust the phase of each dish (as a unit) relative to one dish chosen as the reference dish. However it turns out that this step can, to a large extent, be avoided. Here use is made of the grating lobes from this subarray. Instead of the quite large phase delays that would be needed to steer the subarray to the direction that coincides with the direction in which each dish is pointing, only small phase delays need to be introduced, sufficient to make the *nearest grating lobe* line up with that direction.

4. Attempted Application of the Correlation Array Concept to Sonar

We now turn our attention to two questions. First, can the correlation technique, so useful in radio astronomy, be made to work for sonar? And second, can the technique be harnessed to develop *sparse* 2-D arrays that produce as good a resolution and sidelobe performance as filled (2-D) arrays?

4.1 Assumptions in Radio Astronomy

Let us list what appear to be the relevant differences between the total system in radio astronomy (RA) and the total system in sonar. The relevant assumptions in RA are:

1. The object is in the far field of the telescope. In sonar, this is true sometimes but by no means always.
2. *The signals received from two different parts of the sky are uncorrelated.* In active sonar, the corresponding requirement is given in Section 4.3 below. In general the sonar environment does not conform to this requirement.

3. Assumptions concerning time.
   a. In RA, the image is two-dimensional, having no range component. In *passive* sonar this situation again prevails, often or always. But in *active* sonar, there is a range component; echo ranging is an essential part of the sonar imaging process.[4] The theory applicable in RA needs to be extended to include a range component in the image.
   b. In RA, the time dimension can be used simply for averaging to eliminate uncorrelated components in the signal and thus to improve the signal-to-noise ratio; the time dimension is not already reserved for some other purpose. By contrast, in any sonar system that employs echo ranging, a target at range r is detected essentially by signals received at times very close to $2r/c$. The question needs to be asked whether this use of the time dimension can be combined with the use of time for the purpose of averaging.
   c. In RA, a given correlation run can continue for essentially an indefinite period, limited only by factors such as the rotation of the earth. By contrast, in active sonar, in the case when a short pulse is used, it is suspected that only a correspondingly short time is available for averaging. In sonar systems using a continuous-wave (CW) transmission system instead of a short pulse, this assumption of RA continues to hold. But there is little comfort in this fact, since a CW system cannot measure range with good resolution.

4.2 Near Field

Regarding assumption 1, we consider whether the target's being in the near field precludes the working of a correlation telescope. Two reasons indicate that the near field does not cause any such trouble. First, as shown in Blair [2002], the beam pattern of an array in the near field *is the same as in the far field,*[5] provided that the image point is within the main lobe or the first few sidelobes. This result applies whether the image point is displaced in an angular direction or in the range direction (range sidelobes).

The second reason, not altogether distinct from the first, is that an argument given in Section 4.3 below comes close to deriving Equation (3.5), the Fourier transform relation for a correlation telescope, in the *near* field. The only problem in this argument is that unwanted extra terms (cross terms) are obtained. However, these extra terms have nothing to do with whether the target is in the near or the far field. The above two reasons strongly suggest that the target's *being in the near field* in no way precludes the feasibility of the correlation telescope in the sonar context.

[4] Obviously active probing is impossible in RA—with possible exceptions within the solar system.

[5] Some conditions apply to this result, but they are not very restrictive.

4.3 Correlation of Targets

Encouraged by this 'near field' finding, we now carry out a calculation for a *sonar* correlation array. The system considered is active and the targets are allowed to be in the near field. We derive an equation (for sonar) very similar to Equation (3.5) (for RA).[6] The need, raised in assumption 3a, for an extension from a 2-D to a 3-D image, is thus dealt with.

Recall first that, for a monofrequency sonar system, in the far field, a simple relationship exists between the amplitude beam pattern $B(l,m)$, or response when the array is steered to broadside, and the array grading function $g(x,y)$; this relationship is Equation (3.6), namely

$$\begin{aligned} B(l,m) &= \text{const.}\iint g(x,y)\exp\left[j2\pi\lambda^{-1}(lx+my)\right]dxdy \\ &= \text{const.}\,G(l/\lambda,\, m/\lambda) \end{aligned} \tag{4.1}$$

Here $G(\kappa_x,\kappa_y)$ is the inverse Fourier transform of $g(x,y)$:

$$G(\kappa_x,\kappa_y) = \iint g(x,y)\exp\left[j2\pi(\kappa_x x+\kappa_y y)\right]dxdy \tag{4.2}$$

Equation (4.1) gives the image of a point source or target; the response to many sources is given as a convolution involving (4.1).

We now turn to an active sonar with pulse-echo determination of the range. (The system is no longer monofrequency.) We obtain the three-dimensional image of a collection of point targets. By definition, in (unweighted) delay-and-add beamforming, this image amplitude is obtained *for a general scene* simply by the formula

$$B(r,l,m) = \sum_n E_n\left[\left(r+\left|\mathbf{r}-\mathbf{r}_n\right|\right)/c\right] \tag{4.3}$$

Here $\mathbf{r}=(r,l,m)$, when expressed in essentially spherical coordinates, $\mathbf{r}_n$ is the position of the nth element, and $E_n(t)$ is the signal (voltage) received at that element at time t. For convenience we assume a point transmitter, located at the origin, and we assume the array to be unweighted [as in Blair and Anstee 2000]. For a point target j at $\mathbf{r}_j$, the full near-field treatment [Blair 2002] shows that, under conditions given in that report, the image amplitude is

$$B_1(r,l,m) = C\sum_j a_j\,\xi\left[2c^{-1}(r-r_j)\right]G_1\left[\lambda_c^{-1}(l-l_j,\, m-m_j)\right] \tag{4.4}$$

where for the moment the two subscripts '1' and the $\sum_j$ operator are to be ignored. Here C is a real constant, a_j (a real number) is the target strength of the scatterer j in

[6] Actually, the generalisation of (3.5) to a collection of point targets, given by substituting (3.9) into (3.8).

amplitude terms, the analytic signal $\xi(t)$ is the transmitted signal as a function of t, normalised so that $|\xi(0)|=1$, and λ_c is the wavelength at the central frequency ν_c of the signal $\xi(t)$. ($\xi(t)$ is taken to be a short pulse, but the theory is similar for a coded signal such as a chirp.)

In (4.4), the subscript 1 has been added to B and G to signify that the relationship is written for subarray 1 of a correlation array. The analogous equation holds for subarray 2. For a *collection* of point scatterers, with j referring to the j th scatterer, B_1 is obtained by simply inserting the summation operation $\sum_j$ in Equation (4.4).

A condition on (4.4) is that the bandwidth $\Delta\nu$ of the pulse $\xi(t)$ satisfies $\Delta\nu << \nu$. Despite the 'narrowband' appearance of this relation, we shall call the present treatment a 'wideband' treatment for three reasons: (i) the envelope $\xi(t)$ is explicitly included; (ii) the range resolution due to the bandwidth of $\xi(t)$ is included; and (iii) the treatment is semiquantitatively correct even for $\Delta\nu$ as high as $\frac{1}{3}\nu$.

Now suppose that one multiplies the results for B_1 and the conjugate of B_2 to obtain a combined image.[7] Its intensity is

$$\begin{aligned} B_1(r,l,m)B_2^*(r,l,m) = C^2 \sum_j \sum_{j'} a_j a_{j'}\ \xi\left[2c^{-1}(r-r_j)\right]\xi^*\left[2c^{-1}(r-r_{j'})\right] \\ \times G_1\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_j)\right]G_2^*\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_{j'})\right] \end{aligned} \tag{4.5}$$

where the 2-D vector $\mathbf{l}$ is $\mathbf{l}=(l,m)$ and similarly for $\mathbf{l}_j$ and $\mathbf{l}_{j'}$.

In particular, let us consider a simple case, in which there are two targets, b and c, and their ranges are equal: $r_b = r_c$. Consider the image on the sphere $r = r_b\,(=r_c)$. Because $\xi(0)\xi^*(0)=1$, the ξ factors in (4.5) drop out. We obtain

$$\begin{aligned} B_1B_2^*/C^2 = a_b^2 G_1\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_b)\right]G_2^*\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_b)\right] + a_c^2 G_1\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_c)\right]G_2^*\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_c)\right] \\ + a_b a_c G_1\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_b)\right]G_2^*\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_c)\right] + a_b a_c G_1\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_c)\right]G_2^*\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_b)\right] \end{aligned} \tag{4.6}$$

By the Fourier convolution theorem, each of the first two terms on the right-hand side may be written in terms of Γ, the inverse Fourier transform of the transfer function (see Eqns (3.5) and (3.9)). Thus

$$B_1B_2^*/C^2 = a_b^2\,\Gamma\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_b)\right] + a_c^2\,\Gamma\left[\lambda_c^{-1}(\mathbf{l}-\mathbf{l}_c)\right] + \text{two cross terms} \tag{4.7}$$

Note that for N point scatterers there are N direct terms (terms arising from $j'=j$) and $N(N-1)$ cross terms.

If it were not for the cross terms, Equation (4.7) for sonar would be essentially the same equation as that for the radio telescope, obtained by substituting (3.9) into (3.8). We would thus have an extremely useful sonar instrument—a sparse array without the disadvantages of a normal sparse array—working on the same principle as a cross telescope.

---

[7] Note that the method of combining discussed here differs from that in both the updated and the traditional cross antenna.

It might be hoped that the unwanted cross terms can be shown to be zero—as in fact they can in radio astronomy (as we shall see shortly). But in Section 4.3.1 it will be shown that in active sonar these terms are normally not zero.

In the case where the scene consists of a *single point scatterer*, the system in fact works, because the number of cross terms reduces to zero. This is seen by putting $a_c = 0$ in (4.6). From the surviving term in (4.6), the intensity beam pattern is in fact the product of the beam patterns of the two subarrays (with G_2 conjugated). Being a *product*, the image intensity falls off to small values in *all* directions in an angular distance of λ/L, in contrast to the *sum* pattern (2.2).

To describe the outcome from the cross terms arising when $N > 1$, let us consider again the case where there are just two targets and they are at the same range. Thus (for an image point at that range) Equation (4.6) applies, where the two targets are at (l_b, m_b) and (l_c, m_c). Then, for the cross antenna orientated along the x and y axes, *spurious images appear*[8] *at* (l_b, m_c) *and* (l_c, m_b). These spurious images may be many sidelobes away from the genuine images.

Consider now the case of two targets (j and j') in different range resolution cells. In (4.5), because it contains the product of two ξ s, the cross term between j and j' is very small (governed by the range sidelobes). Thus, in the case of active sonar with pulse-echo determination of the range, the correlation requirement corresponding to assumption 2 of radio astronomy is only that the returns from different directions *in the same range resolution cell* must be uncorrelated.

In radio astronomy, these cross terms are essentially zero. The reason is assumption 2 above (lack of correlation). This assumption holds in RA because *two sources* radiating at a given frequency, even a tiny angular distance apart, will very quickly lose any phase relationship they initially had. They *are uncorrelated* in the sense that their product, averaged over time, approaches zero as the length of the time interval approaches infinity (C&H, pp. 24, 103). As a result, when time-averaging (denoted by angular brackets) is performed, the first cross-term in (4.6) contains the factor[9]

$$\langle a_b a_c^* \rangle$$

which approaches zero. (Each a now represents the strength of a source, not a scattering target.)

[8] This follows because, for example, the third of the four terms becomes essentially $G_1(l - l_b)G_2(m - m_c)$.

[9] The ξ factors drop out because the radio telescope is a monofrequency instrument. All that survives of the waveform ξ is its initial phase. This phase is absorbed into the amplitude factor a_j, making it complex. In (4.5) and (4.6), $a_{j'}$ must be replaced by its complex conjugate.

4.3.1 Conclusions for Sonar; Decorrelation

The corresponding situation in active sonar would be for *the effective source strengths of the point targets to be uncorrelated.* To see what this means, consider a typical cross term in (4.5) when r, r_j and $r_{j'}$ are no longer equal. To evaluate phases, let us drop the two angular factors (G_1 and G_2^*) and, in each factor ξ, consider just the factor $\exp(j2\pi f_c t)$ that represents the carrier wave in $\xi(t)$. The typical cross term is then essentially

$$a_j a_{j'}^* \exp[j(4\pi/\lambda_c)(r_{j'} - r_j)] \tag{4.8}$$

the two terms in r having cancelled out. The phase of this term is what counts in respect of correlation. An alternative argument for (4.8) is to consider the phase change along the go-and-return path from the transmitter to j and back to the array (where presumably the array may be taken at the transmitter for this purpose). The phase *difference* yields the exponential factor. If the target strengths are complex, thus producing a phase change, the factors a_j and $a_{j'}^*$ obviously must be included too. We may define the source strength of target j to be

$$s_j = a_j \exp[-j(4\pi/\lambda_c)r_j]; \tag{4.9}$$

then (4.8) may be written as $s_j s_{j'}^*$. Thus for the cross term to vanish, i.e. for the targets to be uncorrelated, we require

$$\langle s_j s_{j'}^* \rangle = 0 \tag{4.10}$$

where $\langle \; \rangle$ represents some average, to be discussed.

Normally r_j, $r_{j'}$, a_j and $a_{j'}$ are simply constants; there is no averaging and the cross term does not vanish: thus *normally the targets are correlated.* For the targets to be uncorrelated, either a target would have to undergo a random *motion* with a radial component (change of r_j in Eqn 4.9), or else its elastic properties would have to change (change of a_j), during the period of the correlation measurement. Normally neither of these things happens. A second condition is required for lack of correlation, namely, the correlation run must last for a sufficient time for the randomness to be sensed. This requirement is discussed in Section 4.4 and again in Section 11.4.

To make the correlation array work in spite of the correlation problem, it appears that some method for the *decorrelation of targets* must be found. To this end, we may adopt either of two approaches. The first is to persist with the time-based approach, searching for some other feature of the environment that can induce random phase shifts over time. One such method has in fact been used: it makes use of reflections from the disturbed surface of the water. This method is discussed in Section 11.1. It also seems worthwhile to investigate whether fluctuations in the medium give rise to suitable phase shifts.

There is a second approach to achieving decorrelation, which is analogous to synthetic aperture sonar. We shall sometimes call this second approach 'synthetic decorrelation'; however in practice the term 'decorrelation' is commonly used to refer

to the synthetic approach alone. This approach involves setting up (over time) a collection of similar, but not identical, sonar systems. A series of 'snapshots' is taken, one being taken by each member of the collection. (The series is taken in what might be called non-real time.) One then combines these snapshots. This synthetic approach is discussed in Sections 11.2 and 11.3—but in the context of superresolution. It may be possible to adapt some of these methods to correlation arrays. In particular, the forward-backward linear prediction (FBLP) method (Section 11.3) may be of special interest. The method involves using an array that may be regarded as formed from a single subarray moved to various locations. With this method a 'collection' of systems is not needed.

4.4 Time for Averaging

As previously mentioned, averaging over a sufficient time is necessary to remove short-term correlations between targets, even if we are in the very fortunate situation where there is no long-term correlation between those targets. Appendix A considers whether the goal of having sufficient averaging time leads to problems. The argument there suggests that the requirement for sufficient averaging time can be achieved only at the expense of failing to satisfy: (i) the requirement for good range resolution (case of short pulse), or (ii) the requirement for a low level of the range sidelobes (case of a coded pulse). However, this argument ignores the possibility of synthetic decorrelation; see the next Section 4.4.1.

4.4.1 Effect of Decorrelation

In Section 11, various methods for decorrelating the targets are discussed, particularly the FBLP method. The latter does not rely on random phase shifts being induced in the signal in real time. The discussion in Section 11.4 shows that, because of this, *the whole problem of the short time available for averaging goes away* when the FBLP method, or presumably other synthetic processes of decorrelation, are used.

(For overall conclusions on the use of correlation arrays in sonar, see the conclusions given in Section 13.)

5. Introduction to Superresolution

5.1 Initial Introduction; Definitions

Superresolution methods are most often discussed in the context of radio direction finding and, within that, for narrowband signals from sources in the far field. However, some superresolution (SR) work has been done in the areas of radar imaging and sonar; and some SR work has also been done in each of the areas of broadband signals, the near field, and targets as opposed to sources.

Steinberg and Subbaram [1991] (to be referred to as S&S) (p. 303) point out that 'The nominal beamwidth λ/L of an aperture of length L at wavelength λ is not an immutable fact of nature but is a fact of the signal processing of natural objects such as lenses. The operation of a lens ... is governed by diffraction theory ... or approximately by its far-field equivalent, Fourier theory. ...When an aperture is a sampled device such as a phased array, it is no longer fundamental that the ... processing correspond to diffraction theory.' In similar vein, Gething [1991] states (p. 78) that 'In principle there is no limit to the maximum directivity that can be achieved'; however, as we shall see, this statement of Gething's is subject to the existence of an indefinitely high signal-to-noise ratio.

Gething distinguishes between superdirectivity and superresolution. *Superdirectivity* (p. 78) refers to the processing of data from an array to yield the direction of the source or target to greater accuracy than the normal accuracy of around λ/L radians that is obtained by the delay-and-add method of beamforming. (Here L is the length or diameter of the array.) The term also refers to the use of an array to produce a similarly narrow beam when transmitting.[10] On the other hand, *superresolution* refers to the resolving of two targets (or two sources) closely spaced in direction, the quantitative criterion being the same as for superdirectivity.

The techniques of superresolution are actually more general than they appear. S&S (p. 303) say that 'The development of [better-than-λ/L] algorithms and their performance is the subject of modern *spectrum estimation*, as it is called in the time-series literature, or *superresolution*, an apt term in the spatial signal processing or imaging field.' The connection with time series calls for an explanation (given by S&S, p. 308). A time series and its spectrum are Fourier transform pairs. So are the data measured by a linear array and its far-field angular spectrum. Consider the case where the data are produced by discrete sources. With time series, the problem is to estimate the frequency and strength of the (monofrequency) sources. With an array, the problem is to estimate the angular location and strength of the (point) sources. Mathematically the two problems are the same, the difference being simply that one problem is set in the time domain and the other in the angle, or the wave-vector, domain.

In line with the generality just noted, the definition of 'superresolution' given by Gething (p. 198) is an extremely broad one: essentially that it is 'the resolution of multicomponent wave-fields.'[11] Gething adds that an older term for superresolution is '*wavefront analysis*.' The term '*angular spectrum*' is often used in discussions of superresolution.

[10] For a radio antenna used in transmit, Gething (p. 78) quotes an IEEE (The Institute of Electrical and Electronics Engineers) defining standard for 'superdirectivity' as 'the condition that occurs when the ratio of the maximum directivity to the standard directivity obtained with a uniform amplitude equiphase current distribution significantly exceeds unity.'

[11] Gething does not mention the better-than-λ/L criterion at this point; but without it, how could the prefix 'super' be justified?

Superresolution (SR) techniques can be applied in 'many different fields such as radar, sonar, radio astronomy, seismology and spectral analysis, as well as radio direction finding' (Gething, p. 198).

The discussion of SR (Sections 5 to 12) is overwhelmingly on narrowband systems; this is because the literature on wideband systems is quite small. In the text, narrowband conditions are to be assumed unless the discussion implies otherwise. The discussion is rather evenly divided between passive systems (Section 6 and half of Section 8) and active systems (Sections 9, 11, rest of Section 8).

5.2 Literature

Books devoted to SR include Gething [1991], Haykin [1995], Naidu [1996], Naidu [2001], and Stoica and Moses [1997]. Of these, Naidu [1996] is devoted exclusively to the problem of the resolution of frequencies; likewise the Stoica and Moses book has a heavy concentration on that problem as opposed to the problem of resolution of angles. As discussed above, the two problems are closely connected, so that methods that work in the one area are generally applicable in the other. Haykin deals with both problems, while Gething deals with the resolution of angles in the context of radio waves. Naidu's more recent book [Naidu 2001] arrived only recently and will not be discussed in this report, except to comment that it is complementary to his earlier book, being very much concerned with the resolution of angles.

Besides the references given elsewhere in this report, SR (construed to include superdirectivity) has also been discussed as follows. Charland [1989] describes the principles governing superdirective antennas and the factors limiting their realisation. Procopio *et al.* [1964] use what was then a new approach to reexamine in some detail the classical theories of superdirectivity. Rhodes [1971] proves a few results, including that a certain solution is the best in the sense of having the least radiated power in the error pattern. A report by Tucker [1961] has relevance to superdirectivity, while d'Assumpcao and Mountford [1984] have a small segment on the same topic.

5.3 Arrangement of Sections

In Section 6 we discuss SR methods in which the weights applied to the signals from the various elements are fixed, independently of signals received at the array elements. Such arrays may be additive (weighted sum of signals) or multiplicative. In either case, methods with such fixed weights turn out to be not particularly successful. By contrast, weights may be used that depend on the signals, in which case the aim is to optimise the weights in some way. Note that with such weights one necessarily has an *adaptive* array. Various SR methods, based on various optimised weights, are described in Sections 7 to 9. That part of the text is divided into a general introduction (Section 7), methods that estimate a continuous angular spectrum (Section 8) and methods that estimate a discrete angular spectrum (Section 9). Section 10 discusses conditions that

apparently all SR methods must satisfy if they are to work. Section 10 also describes undesirable outcomes that can arise in the image when a method is 'pushed too far.'

For SR, as for correlation arrays, it turns out to be necessary to 'decorrelate' the return signals from the various targets. Methods by which this might be achieved are discussed in Section 11. Section 12 deals with literature on general features of SR, attempting to answer such questions as 'Which method is best?' Conclusions on both SR and correlation arrays are given in Section 13.

6. Superresolution Methods with Fixed Weights: Additive and Multiplicative Arrays

In Section 6 we consider arrays in which the weights are *fixed*, that is, *the weights do not depend on the signals received.* Furthermore in Section 6 each basic weight (in terminology to be developed shortly) is *real,* and steering is achieved simply by applying delays proportional to the path difference. These properties of the weights are the same as for a simple graded array, in which the weights are used to shape the sidelobes; but in SR, the goal is different and the basic weights often oscillate in sign with element position.

To explain the terminology, let us concentrate on additive arrays, in which the image amplitude is simply a weighted sum of delayed versions of the signals at the receiver elements. We shall refer to these weights that do not incorporate a delay as 'basic weights.' In the narrowband case, provided the analytic signal is used, the time delays may be dispensed with; but to compensate for this, the weights must acquire a phase factor $e^{j\alpha}$, which produces the beam steering. These weights that incorporate the delays will be referred to as 'augmented weights.'

The discussion in this section is on systems that are narrowband and passive, except where the contrary is implied.

6.1 Additive Arrays

Superresolution will now be discussed in the context of additive arrays (defined above) with fixed weights.

Urick [1983], when discussing the weighting of arrays, notes (p. 64) that 'an extreme form of shading is called *superdirectivity* ... In a superdirective array, the elements are placed less than one-fourth wavelength apart, with the signs, or polarities, of adjacent elements [i.e. the signs of their basic weights] reversed.'

Gething [1991] speaks more generally of such additive arrays. Speaking of a radio antenna on transmit, he says (p. 78), 'In principle, we can make the pattern of an array of fixed aperture 'sharper' by giving increased [basic] weight to the *higher-order harmonics* in the aperture current distribution. The higher harmonics correspond to rapid reversals in the sign of the current for small movements along the aperture.' To obtain the corresponding statement for a sonar array on *receive,* read 'sign of the

weight' in place of 'sign of the current.' The reference to harmonics applies directly, not only to a circular (circumferential) array, but also to linear or rectangular arrays.[12] Gething's discussion seems to imply that, if D is the period (repetition length) of the sign reversal of the weights, the angular resolution achievable is of order D/L (in place of the usual λ/L). Clearly D cannot be smaller than $2d$, where d is the element spacing (in an equispaced, linear array). Thus first, the resolution cannot be smaller than of order d/L. Second, for SR to be achievable, *the array spacing must be considerably less* than the 'standard' value of $\lambda/2$. Third, the higher the resolution desired, the smaller the spacing d must be. The reader is warned, however, that these three properties do not carry across to SR systems in general and in fact appear to apply only to arrays with fixed weights (see towards the end of this Section 6.1).

Pritchard [1954] showed that, for spacing $d \geq \lambda/2$, the maximum directivity index for a linear (additive) array is produced by *constant shading;* to be precise, the true maximum obtainable differs inappreciably from the value achieved with constant shading. This finding supports the second of the three points just made.

In the context of arrays with fixed weights (at least for additive arrays), superdirectivity is bought at the expense of what might be called low efficiency. Thus, in the words of Gething (p. 78) concerning a transmitting radio antenna, 'In practice, however, it is difficult to obtain superdirectivity. ... Superdirectivity can only be obtained at the expense of a large increase in the ratio of average stored energy to energy radiated per cycle.' For a sonar array on receive, the corresponding statement would read '... at the expense of a large decrease in the ratio of the image intensity, or energy output from the beamforming process, to the energy received in the array.' In other words, the energy output from the beamforming process is considerably less than with a standard delay-and-add beamformer. In other words again, in a superdirective array, there is a *high degree of cancellation* when the signals are combined in the beamformer, *even at the peak* of the beam pattern.

Thus, for the image to appear above the noise, one requires *a high signal-to-noise ratio* in the signals received. Urick [1983, p. 64] makes the same point: 'The price to be paid for superdirectivity is a low array sensitivity—a price usually greater than the gain in directivity index that it buys.' The necessity for a high signal-to-noise ratio is reinforced by a finding of Brown and Rowlands [1959], described in Section 6.3 of the present report ('cannot do better than add the outputs').

Urick notes that the properties of superdirective arrays were pointed out by Pritchard [1953]. Furthermore, Urick (p. 62) notes, 'the selection of shading factors to achieve a desired beam pattern can be done by the method of Lagrangian multipliers, as was pointed out originally in a paper by Pritchard [Pritchard 1954].'

It is appropriate at this stage to review two trends that have emerged from fixed-weight, additive systems, and, in respect of at least one of these trends, to caution against the tendency to generalise from these systems to other SR systems. The first trend is the conclusion suggested by the descriptions of Gething and Urick that, for SR

[12] Gething (p. 79) gives references that apply the concept of spatial harmonics to other arrays, including systems with non-linear processing of the element signals.

to be achieved, the array spacing d must be considerably less than $\lambda/2$. Two pieces of evidence show that *this result does not extend to adaptive arrays.* First, R. Niland [private communication] reports that the Capon method (a widely used, adaptive SR method) does not have this requirement. Second, Steinberg and Subbaram [1991] (reported in Section 11.3 of the present report) describe an adaptive SR system in which $d = \lambda/2$, in contradiction to the 'requirement' $d << \lambda/2$. It is therefore believed that the $d << \lambda/2$ requirement applies only to fixed-weight systems.

We now turn to the second trend, namely the finding that the efficiency is low and that the required signal-to-noise ratio is therefore high. *This finding does appear to extend to adaptive arrays.* R. Niland [private communication] reports his belief that this is so, at least for Capon's method. This belief is based in part on a pattern that is often seen in the Capon weights when the array is steered to broadside. In a typical case, across the total of about 100 elements, the (real) weights showed a strong tendency to oscillate (with position) about a mean near zero, with a period of about 10 elements. This indicates a low efficiency and the need for a high SNR. The evidence *in this particular example* is overwhelming; the result may extend to adaptive SR in general.

By way of fine-tuning to this discussion, the following is mentioned. The emphatic language of Urick and Gething gives a hint that fixed-weight arrays perform *worse* than adaptive arrays in regard to SNR (requiring an even higher SNR), but we leave that as an open question.

6.2 Interlude

Brown and Rowlands [1959] took Shannon's law of information flow along a channel and applied it to arrays. They showed that for $S/N > 1$ (where S/N is the signal-to-noise ratio at a single element), one *can* do better in gaining information than using delay-and-add combined with a constant spacing. Two possible ways are: (i) the use of element spacings following a geometric progression, and (ii) the use of nonlinear operations, in particular multiplicative arrays (discussed in Section 6.3). Related work of Brown and Rowlands is described in Section 6.3.

6.3 Multiplicative Arrays

In multiplicative arrays, signals are multiplied as well as added. These arrays are discussed by Urick [1983, p. 65] and Gething [1991, p. 89]. Urick notes that, in some proposed schemes, products of as many as three or more voltages are formed.

The discussion of Gething is based largely on a paper by Shaw and Davies [1964]. By way of illustration, a linear array of eight equispaced elements is considered. The array is divided into two halves and the resultant signals (resultant equals sum of the four voltages) are multiplied together. The resulting beam pattern (with no further squaring to convert 'amplitude' to 'intensity') was compared with that of the unweighted additive array, where in the latter case the amplitude beam pattern is

squared. The multiplicative beamwidth turns out to be about 70% of the beamwidth of the square-law-rectified additive pattern—a clear improvement.

However, this is not the end of the matter. For multiplicative arrays, the response to a *pair* of sources *cannot* be inferred from the response to a single source, and requires a further calculation. It turns out that three effects can vitiate the superiority of the multiplicative system as follows.

1. Cross product terms are obtained, similar to terms obtained in the correlation telescope in radio astronomy, discussed in Section 4. For the purpose of a good image without artefacts, it is almost certainly necessary that, as a result of an average over time, these terms vanish. But, as discussed by Gething [1991], the time-averages of the present cross product terms, like those in the sonar correlation array, do not vanish if there is *coherence* (correlation) between the two sources (or between the two targets in the active sonar case). As discussed in Section 4.3, in the active sonar case, coherence is virtually unavoidable. Thus the multiplicative array fails—unless some method of decorrelation (Sections 4.3.1 and 11) can be applied.
2. When the two targets are of unequal strength, with an amplitude ratio differing substantially from unity, the multiplicative system tends to suppress evidence of the weaker signal. This does not bode well for the multiplicative array if one is aiming to see structure in the mine image.
3. For the device to work, the signal-to-noise ratio must be high. Thus Urick states that 'The saving in number of array elements is coupled with a degradation of the signal-to-noise ratio [of the *output of the array*] when the signal-to-noise ratio [of a single element] is less than unity.'

In regard to effect 3, Brown and Rowlands [1959] considered the situation where one has n elements and these can be regarded as independent of each other. They showed that, when S/N (i.e. at a single element) is small compared to one, one *cannot do better* than *add the outputs* of the n elements. (They seem to assume an equispaced array, but do not explicitly state this.) This finding is further evidence for the effect 3.

In related work, Brown and Rowlands [1959] compared the accuracies of two common methods for effectively rotating the array to find the direction of a source, both for high S/N and for low S/N.

Gething (p. 92) mentions sonar, saying, '[Multiplicative processing] is also used in sonar, sometimes combined with hard-limiting (Nairn 1968).' Actually the paper of Nairn does not seem to contain any multiplicative processing. It does however contain hard-limiting: the outputs of the individual transducers were replaced by 0 or 1 depending on whether a certain threshold was exceeded. Thus it may be said that nonlinearity (if not multiplicative processing) is a feature of this beamforming of Nairn.

Urick summarises: 'it may be said that multiplicative arrays find applications in conditions of high signal-to-noise ratio where narrow beams or high resolution are desirable or where a reduction in size or number of elements over a corresponding linear array is mandatory.' He goes on to mention complexity as a problem. However, Urick's summary does not tell the full story. He fails to mention the important problem that arises when there is coherence between the returns from different targets or sources.

6.4 Correlation Arrays

These arrays in essence involve forming the element-by-element products of the voltages of two arrays, and integrating each product over time. They may therefore be viewed as a form of multiplicative array. Correlation arrays have been discussed in Sections 2 to 4, where the problems in the sonar context have been pointed out.

On a positive note, both Clarke [1970] and Shearman *et al.* [1973] succeeded in forming simple images, in the areas of sonar and radar respectively, each using a variant of the correlation array (see Section 11.1).

7. Superresolution Methods with Adaptive Weights: A. General

In the methods described so far, no rationale is given for selecting a particular set of weights (with the possible exception of Pritchard's work). In the methods to be discussed in Sections 7 to 9, a rational procedure is given by which the weights are optimised in some sense. An essential step in the procedures to be discussed is to make the weights depend on the signals received; thus the weights are *adaptive.* The use of adaptivity appears to lead to a marked improvement in performance over fixed-weight arrays in one area (the ratio of the spacing to $\lambda/2$), as discussed towards the end of Section 6.1. (Adaptivity possibly also leads to improvement in the area of signal-to-noise ratio.) Because of this marked difference, when we speak of 'superresolution' (SR) in the remainder of this report, we *exclude* fixed-weight arrays except where the context implies otherwise.

A consequence of the adaptive feature is as follows. Consider the narrowband, far-field case. In fixed-weight, additive beamforming, when the 'look' direction is changed, the absolute values of the (augmented) weights remain unaltered, and their phases are altered in proportion to the changes in the path length. But with SR beamforming, when the look direction is changed, neither of these properties holds.

Sections 7 to 9 form a group. The methods of Section 8 result in an estimate of a continuous angular spectrum, while those in Section 9 produce a discrete angular spectrum. But there is a further difference as follows. Each method in Section 8 simply involves a selection of weights followed by beamforming with those weights. The methods of Section 9 involve a further level of interpretation, in which the data are attributed to a finite number of point sources or point targets. The problem is to estimate the number of targets, their position and their strength.

Nash [1994, pp. 27–42] outlines the six main SR spectral estimators. These are said to be Linear Prediction (LP), Maximum Entropy Method (MEM), Minimum Variance (MV) (also called Capon's method), Prony's Method (PM), Multiple Signal Classification (MUSIC) and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT). Nash goes on to discuss three of these (LP, MV and MUSIC) in more detail in the context of the Inverse Synthetic Aperture Radar (ISAR) problem.

More detailed references, not given below, are: Makhoul [1975] for LP, Marple [1987] for MEM and MUSIC, and Kay and Marple [1981] for PM.

Discussions in the book by Steinberg and Subbaram [1991] are worth paying particular attention to, because it is concerned with imaging—albeit microwave imaging— rather than detecting sources, in other words it is concerned with active systems. In addition, it is concerned largely with complicated targets rather than points.

8. Superresolution Methods with Adaptive Weights: B. Estimation of Continuous Angular Spectrum

8.1 Capon's Method and Related Matters

A colleague has suggested that Capon's method would be a good starting point for applying SR to the Minefinder project. The method gives a rational procedure for choosing and indeed optimising the weights of an additive array. For a narrowband detector, the absolute values and phases of the augmented weights chosen depend on the set of complex voltages received (and on the steering direction).

In this Section 8.1, the discussion up to the end of Section 8.1.1 is on systems that are narrowband and passive. The remainder of Section 8.1, however, contains much that is wideband material and much that is on active systems.

8.1.1 Capon's Method

Capon's method [Capon 1969] is described in Haykin [1995, pp. 14–17] and briefly in Steinberg and Subbaram [1991] (S&S) (p. 320), both in the context of sensor arrays. It is also described in Naidu [1996, p. 244] in the context of time series, i.e. without reference to arrays. The method has alternative names: it is called by Haykin the *minimum variance* (MV) technique, and by S&S the *maximum likelihood method* (MLM). Nash [1994, p. 4] throws light on these names, noting that maximum likelihood is the original name given by Capon, but that the name is a misnomer, so that Capon's method is now generally referred to as the minimum variance method.

Haykin begins by describing the usual delay-and-sum technique, but in matrix notation, for a passive, narrowband system detecting incoming signals over time. Augmented weights are applied that contain a phase factor $e^{j\alpha}$, where α is proportional to the time delay needed to steer the array to the desired angle θ; the sequence of these weights (over the elements) forms the weighting vector $\mathbf{w}(\theta)$. Haykin concludes that the delay-and-sum technique is suitable for a *single* source; but not for multiple sources. The latter assertion is perhaps surprising, as one might have thought that the technique, though of not-very-high resolution, is at least robust. But he goes on to explain the two problems that arise with multiple sources: 'However, if more than one source is present, the peaks may be *at the wrong directions* or *may not be resolved at all* because of the relatively poor resolution properties of this technique'

(italics added). In this last clause he implies that it is not difficult to achieve considerably greater resolution than that given by delay-and-add; that is, it is not difficult to do this in the context of radio direction finding.

Haykin then proceeds to the Capon method. 'Capon realised that the poor resolution of the delay-and-add technique can be attributed to the fact that the power of the delay-and-sum processor at a given direction does not depend only on the power of the source at that direction, but also on undesirable contributions from other sources.' Capon's approach consists of two parts. First, the array is still steered to a direction θ via a linear combination of the received signals, using a weighting vector $\mathbf{w}(\theta)$ (vector of augmented weights). However, each element of $\mathbf{w}(\theta)$ now contains an *additional* phase factor $e^{j\beta}$ (and also an additional real factor).

Prior to the optimisation step, a 'steering condition' is invoked that puts a linear constraint on $\mathbf{w}(\theta)$. Though stated by Haykin, this constraint is better expressed by S&S as follows: the constraint is that *the array gain is unity at the look direction* θ.

At this point we introduce terminology from Haykin as follows. At the look angle θ, first there is the contribution from the desired target or source, which is very close to θ. The remaining contribution to the received signal, called *interference,* is due to the other signals entering through the sidelobes, together with noise.

Haykin moves on to the second part of Capon's approach, as follows. 'Now in order to minimise the contribution to the output power of other sources at directions different from θ, Capon proposed to select the vector $\mathbf{w}(\theta)$ so as to *minimise the output power'* (italics added) *subject to the steering constraint.* Here the output power concerned is the power received at the look direction θ. One can make intuitive sense of this minimisation principle as follows. The constraint fixes the output power received at θ *due to the desired signal.* So (provided powers are additive) the minimising must minimise the power received at θ due to the interfering signals and noise. This is just what one wants to do.

Note that the 'intuitive sense' argument does not cover the case where an interfering signal is *correlated* with the desired signal (since powers are then not additive). Haykin comments: '[This] minimisation of the power output ... may create serious difficulties when the interfering signals are correlated with the desired signals. In this case the resulting vector [$\mathbf{w}(\theta)$] will operate on the interfering signals so as to partially or fully cancel the desired signals, as dictated by the minimisation requirement.' The possible existence of these 'serious difficulties' associated with correlation suggests that, if an *active* system is used in conjunction with the Capon method, some method of decorrelation (Sections 4.3.1, 11) will be essential.

After the optimum $\mathbf{w}(\theta)$ has been calculated by the two parts of Capon's approach as above, the angular power spectrum is obtained by inserting that value into the usual beamforming formula for power. The result should be equal to the true (source) power spectrum, combined with *less smoothing* that with delay-and-add; in other words the result should be better resolution.

Haykin proceeds to describe an alternative version of the minimisation principle, which is equivalent to the above one *provided that the interfering signals are uncorrelated with the desired signal.* The alternative is that 'the minimum variance technique can be

interpreted also as *maximising the signal-to-interference ratio* at the beamformer output' (italics added). Here 'interference' has the technical meaning given above and the same 'steering constraint' as before is applied. This new version has more intuitive appeal.

The description of Capon's method that S&S give is short and will be discussed only briefly. The formula they quote for the angular power spectrum is essentially the same as that of Haykin, except in one respect: the formula (11-19b) in S&S for the correlation matrix has a sharp cut-off imposed at an integer P (where P is defined in Section 9.1). The description turns out to be of the Capon method but with two restrictions imposed as follows. First, it is *assumed* that the interfering signals are uncorrelated with the desired signal. Second, the overall system is modelled using an autoregressive process of order P (see Section 9.1).

8.1.2 Broadband Signals; Superresolution in Range; Suppression of Sidelobes

Three items from the literature are mentioned here that may have a particular bearing on the application of the Capon method (or SR more generally) to the Minefinder problem. The items are, first, work on broadband signals (by Nunn); second, work on superresolution in range (by Nash); and finally, work on the suppression of sidelobes (by Ferguson).

Further work on broadband systems is relevant in its own right. Contributions in this area have included Su and Morf [1983], Wang and Kaveh [1985] and Hung and Kaveh [1988].

Nunn [1985] considers the problem of an adaptive array in the presence of broadband sources. He first reviews items of earlier work that are based on either the frequency domain approach or the time domain approach. A difficulty with the former is that, for a full frequency domain approach, the computational load is (at the time of his writing) enormous. The latter approach suffers from being highly suboptimal, but is more robust than the frequency approach. Nunn obtained 'very good' performance by developing what he called the Suboptimal Frequency Domain Processor (SFDP).

We turn to pertinent comments made by Nash [1994, pp. 6–10]. Nash distinguishes between SR in cross-range, SR in range, and SR simultaneously in cross-range and range. SR in cross-range 'involves generating the complex range profiles using conventional processing, then applying a 1-D [SR] spectral estimator down each range bin to achieve [SR] in cross range.' For SR in range only, much depends on whether stepped frequency waveforms or chirp waveforms have been used. 'For stepped frequency waveforms exactly the same algorithms [as for] cross range can be used. ... 1-D Fourier transforms would be performed across the spectra of the range profiles resulting in cross range profiles, the 1-D [SR] spectral estimates would be formed across the cross range profiles.' Nash proceeds to briefly describe options for the chirp waveform.

Nash describes a number of techniques to increase the resolution simultaneously in range and cross range dimensions (the third type of SR). There are two basic approaches: use 1-D SR techniques in both dimensions or use 2-D SR techniques. The

latter approach is preferable, as it requires fewer parameters and so 'the parameters from the 2-D [SR] technique will be more accurate.' Some references given by Nash appear to be quite relevant.

Ferguson [1998] has applied the Capon beamformer (Minimum Variance Distortionless Response beamformer) to the case where one calculates a 2-D spectrum, which in this context refers to the spectral intensity as a function of both frequency and wave number. He shows, for a towed, low-frequency sonar, that as well as improving the spatial resolution of the array, the method suppresses sidelobes in a case where the latter are very high.

8.1.3 Thoughts on Application of Capon's Method to the Minefinder System

Early in the preparation of this report, some thought was given to how the Capon method might be generalised to deal with the Minefinder system. While these thoughts have been overtaken by later literature survey work (particularly methods described by S&S), they are included as Appendix B, in case some of them may prove of use in the Minefinder problem.

8.2 Maximum Entropy Method (MEM)

About the same time as Capon's method was developed, Burg [1967, 1972] developed the maximum entropy spectral estimate (or maximum entropy method—MEM). The rationale for this method is that one chooses the spectral estimate that makes the least number of assumptions about the data; that is, one adds in the least amount of spurious information. The MEM method has been discussed by S&S [1991, p. 320], Gething [1991, p. 204] and Nash [1994, pp. 4, 28].

8.3 Model-Based versus Non-Model-Based Methods

Both Haykin [1995, Chapter 1] and S&S (p. 320) emphasise this distinction. Model-based methods assume 'structural and statistical models of the signals and noise' (Haykin, p. 18). For example, the FBLP method (Section 11.3) assumes, at Equation (11-12) of S&S, an autoregressive moving average (ARMA) process or model (in addition to assuming other models). To motivate this model it is noted that, for a scene of T point targets (in the far field), in the absence of noise, the signal at the nth element can be predicted exactly as a weighted sum of the signals from the T elements lying immediately to the left. The ARMA model assumes that this relation holds also in the case where there is noise. This assumption makes ARMA into a model rather than known fact.

Non-model-based methods include delay-and-sum and the Capon method. Model-based methods include the FBLP method and the optimal solution of the DSWN (Deterministic Signals and White Noise) model (Haykin, pp. 25–33). Regarding

MUSIC (Section 9.3), S&S state that it is non-model-based, while Haykin (p. 23) appears to imply that it model-based.

9. Superresolution Methods with Adaptive Weights: C. Estimation of Discrete Angular Spectrum

As mentioned in Section 7, these 'discrete' methods involve modelling in which the data are attributed to a finite number of point sources or point targets. The problem is to estimate the number of targets, their position, their strength and perhaps their frequency.

The discussion, both in this section and in the later Section 11, is based largely on the work of Steinberg and Subbaram [1991]. Their description is for systems that are narrowband but that, fortunately for the Minefinder application, are also active.

9.1 An ARMA Model

Steinberg and Subbaram (pp. 304–319) give a discussion that homes in, in stages, on a decorrelation method called the forward-backward linear prediction (FBLP) model. In this Section 9.1 we trace the beginning of this thread. (The remainder of this thread will be traced in Section 11.)

S&S begin by considering all-pole, or autoregressive, methods. They then specialise to an autoregressive moving average (ARMA) model, which they solve. The present Section 9.1 will progress to that point. The discussion by S&S starts by treating targets as *sources*; later S&S take account of the fact that the effective phase of each of these 'sources' depends on the transmitter-target distance.

The all-pole methods are motivated as follows. For an equispaced array of spacing d, let $\hat{s}(z)$ be the estimate of the source distribution, where $z = \exp(-jkud)$, $k = 2\pi/\lambda$, $u = \sin\theta$ and θ is the angle from broadside. It is pointed out that, for the simple delay-and-add method, $\hat{s}(z)$ is a polynomial in z of degree $N-1$, where N is the number of elements. Thus $\hat{s}(z)$ has zeros but no poles. 'The ... zeros of $\hat{s}(z)$ locate the nulls and deep dips in the real image ... Of primary interest to the user, however, are not the nulls in the pattern but the peaks, for they disclose the directions of arrivals of the energy and are bearings to the target' (S&S). Due to the absence of poles (near the unit circle of z), the peaks are broad. By contrast $\hat{s}(z)$ is, in a sense, sharp near the zeros. Instead of the zeros of $\hat{s}(z)$ that we have at this stage, we would like to have zeros in the *denominator* of $\hat{s}(z)$, and have them located at the target points. This would yield sharp peaks.

We therefore seek a new estimator $\hat{s}(z)$ of the angular spectrum that resembles the reciprocal of the delay-and-add $\hat{s}(z)$, but with the augmented weights chosen differently. They are to be chosen so that the zeros of the denominator occur at the

target locations. S&S discuss *all-pole models,* also called *autoregressive models;* these are models in which $\hat{s}(z)$ has poles but no zeros.

S&S (Section 11-2) then consider the problem of finding the weights within an *autoregressive moving average* (ARMA) *model.* They give a solution using the least-squares method. In the process, a couple of models (approximations) are adopted which together constitute the overall model. First, a moving-average model is adopted, which draws on the analogy of an array to a time series. As a result, $\hat{s}(z)$ is obtained as the ratio of two polynomials. A second model is applied, which makes the numerator reduce to a constant. This yields the desired all-pole solution (their Equations 11-16 and 11-19a). In the case where there is no noise—i.e. the case where the assumed finite number of targets exactly represents the data—the poles lie exactly at the targets.

At the beginning of their least-squares solution to the ARMA-model problem, S&S introduce the assumption that the targets can be modelled by a finite number, T, of point targets. For this reason (discreteness), the discussion of all-pole methods has been included here under Section 9. It appears that S&S do not discuss how to estimate T; they do however discuss relations between N, P and T required for the set of equations to be well-conditioned. Here N is the number of elements and P is the number of element voltages from which the next voltage in line is assumed to be predictable; later in the development, subarrays of size $P+1$ are formed. P is called the *order* of the model.

The above least-squares solution is subject to an important proviso. The least-squares method, being quadratic, requires among its inputs the expectation value

$$R(m,n) = E\{e_m e_n^*\} \tag{9.1}$$

where e_n is the voltage at the nth element. In this context the expectation value (denoted by E) means (in the passive case) the average over an ensemble in which the sequence $(\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_s)$ of the phases of the source signals takes on all possible values. Thus every relative phase $\alpha_i - \alpha_j$ is spread uniformly over the interval $(0, 2\pi)$. (For a target as opposed to a source, the 'phase' is defined below Eqn (4.8).)

For *incoherent* sources (sources whose phases wander in time with respect to each other), the above expectation value is relatively simple to obtain experimentally: one simply averages $e_m e_n^*$ over a sufficiently long period of time; that is, one takes a sequence of snapshots and then averages (S&S, p. 311). For sources or targets that are *coherent* with each other—and this includes essentially *all targets*—the time-averaging method fails. It is possible to get around the problem by means of special experimental arrangements, as is discussed in Section 11.

9.2 Estimating the Number of Sources or Targets

Only the passive case will be discussed, as the literature primarily deals with this case. Haykin (pp. 6, 21–22) describes a method for estimating the number of sources, based on the Minimum Description Length (MDL) principle [Rissanen 1989]. The estimation

of the number of sources via MDL is a preliminary procedure to be carried out prior to performing a further procedure (such as MUSIC, Section 9.3) to estimate the *locations* of the sources.

We briefly describe the MDL method. Suppose that there are q narrowband sources. q is estimated [Haykin 1995, pp. 21–22] by considering the sample covariance matrix, which is essentially the set of products of the received voltages. Provided both the 'White Noise' model and the 'Stochastic Signals' model hold (Haykin, pp. 3, 18), and subject to a nonsingularity condition, the $p-q$ smallest eigenvalues of the covariance matrix are all equal. The method for obtaining the estimate $\hat{q}$ of q is based on this fact.

It is not clear whether the present method can be used in conjunction with the ARMA model and FBLP method (Sections 9.1, 11.3). It appears that the approaches do not mesh closely; this suggests that the methods cannot be used together.

9.3 The MUSIC Estimator

This estimator has been mentioned by a colleague as relevant to sonar. It is described by Haykin [1995, pp. 23, 226 *et seq.*], by Gething [1991, e.g. pp. 207–208] and by S&S (pp. 320–321). Gething introduces the estimator thus: 'MUSIC is an acronym for Multiple Signal Characterisation. The MUSIC procedure, proposed by Schmidt [1979, republished 1986], is now easily the best-known and most frequently-cited method of superresolution.'

Nash [1994, pp. 4–5] gives an interesting, brief historical sketch of developments in SR. Included is the comment that, following the Capon method, 'the next major step in [SR] spectral estimation was by Pisarenko [Pisarenko 1973] who used eigen decomposition techniques to form Pisarenko's method. [That] method was then generalised by Schmidt ... to form the Multiple Signal Classification (MUSIC) algorithm.'

Our brief discussion of MUSIC will follow that of Haykin (p. 23). Consider the case where an array of p elements is used in a situation where there are q narrowband sources, each of frequency close to ν_0. It is desired to estimate the number of sources, their angular location and their strength. As a preliminary to MUSIC, q is estimated as in Section 9.2.

The procedure for MUSIC is as follows (Haykin, p. 23). A formula for an estimator $g(\theta)$ is written down, which, if all the estimates leading up to it were exact, would be equal to zero at the location θ of each source. One then selects 'the $\hat{q}$ values of θ that yield the lowest minima.' (Presumably the selection is made from amongst all the local minima.)

Nash [1994, Chapter 6] conducted an experiment to compare three methods of analysis in the context of ISAR: LP, MV and MUSIC. '... the MUSIC estimate [was] shown to be the best [SR] technique of the three tested ..., and was shown to be a superior and practical alternative to the conventional estimate ..., for the case of a small number of observations.'

9.4 Other Estimators

Among other methods for estimating the discrete angular spectrum we mention ESPRIT [Gething 1991, p. 208–210; Naidu 1996, pp. 313, 380]. The original ESPRIT, due to Paulraj *et al.* [1986], is described by Gething. Gething makes the significant comment that 'ESPRIT, like MUSIC, encounters difficulties in resolving coherent rays' and mentions methods that have been used to overcome this problem.

Haykin [1995, pp. 172, 185] discusses some variants of the standard ESPRIT method, mainly UCA-ESPRIT, where UCA stands for 'uniform circular array.' He says that 'UCA-ESPRIT is a novel algorithm that represents a significant advance in the area of 2D angle estimation. It is a *closed-form* algorithm that provides *automatically paired* azimuth and elevation estimates for each source' (italics in original). Here the term 'closed-form' indicates that an iterative search procedure is not required; the algorithm provides directly-calculable estimates of the source directions.

Other algorithms related to MUSIC are discussed in Gething (pp. 210–213) and in Haykin (Chapters 5, 6).

10. Possible Drawbacks of Superresolution

This section discusses possible drawbacks of SR. These include both undesirable consequences of using SR when the situation is unfavourable, and requirements that must be fulfilled if SR is to work. Such drawbacks are as follows.

1. As discussed at the end of Section 6.1, a *high signal-to-noise ratio* is thought to be necessary for the successful application of SR. Ideal conditions for a high SNR often occur when there are just a few point sources (or targets), as in radio direction finding. (This is because there is little interference due to the signals from the other sources.) In the Minefinder sonar problem, some comments are appropriate. First, in the situation of mine detection at the *longest* ranges, where the signal is barely seen above the noise, the use of SR seems inadvisable. SR should, however, be good for detection and classification at the closer ranges. However even here, the question must be raised as to whether interference from other point targets in the scene (on other parts of the minelike object, for example, in the case of classification) produces sufficient 'noise' to spoil the image.
2. *Correlation* (coherence) between the returns from two targets can perturb the angular spectrum, as discussed in Sections 8.1 and 9.1. Artefacts are produced unless some method of decorrelation is found. In the case of active sonar, in which range as well as direction is estimated, it is believed that such effects occur only due to pairs of targets that lie within the same range resolution cell (but different directional cells). On the other hand in active sonar, correlation occurs between virtually all pairs of same-range targets. Thus in the Minefinder situation (prior to any decorrelation), SR faces a problem due to correlated targets. (This problem would presumably go away if there were but

a few targets, randomly placed in range and angle.) Fortunately, however, methods of decorrelating targets exist; these are discussed in[13] Section 11.

To the above drawbacks should be added the problem that relatively little work has been done on applying SR to wideband systems. In addition, active systems are very underrepresented in the literature.

An over-zealous attempt to increase the resolution can lead to unfortunate outcomes. In detail, these are as follows.

1. A true target is detected, but in the wrong direction. (Note that this occurs already, to some extent, with standard delay-and-add (Section 8.1.1)).
2. False extra targets may be detected. For essentially the reasons given in Section 4, this is the likely outcome when there are coherent targets.
3. A peak may break up into two peaks by the formation of a null or minimum within it.

Evidence for outcomes 1 to 3 is contained in a passage from S&S (p. 317) concerning one of the 'four' methods mentioned in Section 11.3 (not the favoured fourth method). They say, 'The SR images ... also exhibit spurious peaks [outcome 2], biases in the target location estimates [outcome 1], and peak splitting [outcome 3].' Actually, the favoured fourth method also gave false peaks for, as S&S (p. 327) say, '... some false targets are introduced due to the large model order and the low SNR per element.' This last result suggests that, in sonar, while one should try to minimise the number of false peaks, one may have to live with a certain number of them.

This last example (the fourth method) also throws light on the question, 'Where do the extra peaks occur?' The separations between peaks can be estimated from Figure 11-5(c) of S&S. Typically the false peaks are seen at 6–10 SR beamwidths from the genuine targets, where the SR beamwidth is approximately 4 times smaller than the normal delay-and-add beamwidth. The figure of '6–10' is large; thus *the effect is serious in relation to the resolution aimed for.*

Interestingly, for a multiplicative array with fixed weights, Figure 5.6 of Gething [1991] shows how simply changing the phase difference between two sources can produce a dramatic effect on the angular response. This may throw some light on, for example, outcome 3 for a SR array.

11. Methods of Decorrelating the Targets

As mentioned in Section 9.1, the ARMA method succeeds only if the targets are incoherent with one another. With active systems, this condition is almost never

[13] For balance it should be pointed out that the simple delay-and-add system, i.e. the 'competing system,' is not entirely immune from interference effects. Thus Haykin (quoted in Section 8.1.1) states that with delay-and-add, the presence of two or more sources can cause the peaks to be in the wrong directions. However, as a few examples show, the shift in the local maximum is at most by a fraction of a beamwidth. Thus the effect is small, at least on the scale of the resolution one is trying to achieve with delay-and-add. (It is not small on the scale of the resolution that an SR method aims for.)

satisfied. Yet with special experimental arrangements, effective incoherence can be produced, as discussed by Steinberg and Subbaram [1991, Sections 11-3, 11-4, Chapters 5, 6].

S&S (Sections 11-7, 11-3, 11-4) list four methods of achieving decorrelation as follows. The first method is to sequentially use transmitters at different locations; this is called *transmitter-location diversity*. The second is to physically move the receiving array to different locations; again the array locations are used sequentially. The third method is to arrange for the target to be moved. The fourth method requires only a single pulse from a single transmitter. It relies on a method of signal processing that requires the receiving array to be periodic. The situation is *as if* the receiving array had moved. Before discussing methods typified by the four above (Sections 11.2, 11.3), we consider a group of miscellaneous methods of decorrelation (Section 11.1).

11.1 Methods Based on Real-Time Averaging

We discuss here two pieces of work, first mentioned in Section 6.4. The first is that of Shearman, Bickerstaff and Fotiades [1973] (SBF). SBF discuss over-the-horizon radar using reflections from the ionosphere. They used a two-antenna receiving instrument, with variable separation. Since also the two outputs were multiplied and the result averaged over time, the device was in essence a correlation array (see Sections 2 to 4), but with an extra twist (see next footnote).

Prior to the experiment, SBF thought that 'the *random variations in the height of the ionosphere* would decorrelate the returns sufficiently' (italics added). This additional method of decorrelation (besides the methods of varying range and varying target strength, suggested in Section 4.3.1) is worthy of note. The discussion of SBF (p. 416) makes clear that the averaging is to be done in *real time*.[14] This feature contrasts with the *synthetic* averaging approach to be developed in Sections 11.2 and 11.3. In the SBF experiment, it turned out that the returns were not decorrelated sufficiently.

The response of SBF was that 'the radar was modified to make the transmitter and receiver *frequency-hop* from pulse to pulse randomly ... This would serve to decorrelate echoes from targets differing in oblique range by a few km' (italics added). (Presumably, 'differing in oblique range' refers to targets at the same range but at different bearings.) The principle that makes the frequency-hop method work is not clear from their description. However the method has similarities with the frequency-diversity method described in Section 11.2. The frequency-hop method appears to be based, at least in part, on synthetic decorrelation as opposed to real-time averaging. At the time of writing by SBF, encouraging experimental results had been obtained.

Thus SBF have proposed two methods. The first (the use of the random height) suggests that in sonar, one might rely simply on fluctuations in properties of the water such as the temperature to produce decorrelation (see also the work of Clarke below). The second method, frequency hopping, might be copied across directly.

[14] To complicate matters, there is also a 'synthetic' component to the experiment. As the experiment proceeds, of the only two elements, one element is gradually moved away from the other. Thus a synthetic aperture is produced.

The above work by SBF was based on earlier work by Clarke [1970]. Clarke developed theoretically a proposed method for locating radar sources in the presence of a moving ionosphere. To test the principles, he performed a sonar experiment. Regarding the radar method, essentially this method was later employed by SBF as their first method, i.e. prior to their use of frequency hopping. We concentrate here on the sonar experiment. Clarke's work is described by SBF as follows: 'the sky-wave backscatter situation was modelled using sonar under water, reflection from the ionosphere being simulated by reflection of upgoing ultrasonic waves downwards from the rippled surface of the water.' The sonar experiment was said to be successful in achieving decorrelation (and in resolving sources) (Clarke, Section 9). Clarke also studied experimentally the cross terms due to the correlation of sources by performing an experiment without the surface reflection.

Thus, for the Minefinder system, another possibility for producing decorrelation is to have the acoustic waves reflected from the surface with its irregularities. Note however that with this method the operator is forced to work with whatever sea state happens to exist.

11.2 Diversity Methods

We briefly discuss some decorrelation methods based on the concept of 'diversity.' The methods discussed are ones described by S&S, mentioned at the start of Section 11.

Regarding the first method (transmitter-location diversity), while S&S mention it as a possibility for use as a decorrelation method within SR (p. 307), their detailed discussion of this method (S&S, Chapter 6 and Section 5-8) is as *a means of improving the distant sidelobes produced by a sparse, random array.*[15] The present discussion likewise is limited to this application. The details are given in Appendix C. Here two comments are worth making. First, the transmitter-location diversity method exists in an 'incoherent' and a 'coherent' version. Speaking of the 'incoherent' method, S&R (p. 121) point out that 'One importance of this procedure is that the image processing tools developed in radio astronomy now become applicable to terrestrial microwave imaging.' Second, their use of the term 'decohering' (p. 121) in describing the transmitter-location diversity procedure suggests again that there is a useful connection with superresolution.

The second method of achieving decorrelation is to physically move the receiving array to different locations without any internal rearrangement of the elements[16] (S&S, pp. 312–313). This arrangement is called *'aspect-angle diversity'* (S&S, Section 5-7). The procedure can be carried out, for example, with airborne radar arrays. S&S quote an

[15] S&S (Chapter 5) also discuss the use of 'diversity' of other kinds as a means of improving the sidelobes produced by a sparse, random array—for example frequency diversity.

[16] In a generally excellent book, page 312 of S&S is unfortunately written in a quite confusing way.

example [Giuli 1986] in which this second method was used to obtain *superresolution* images.[17]

A particular case of aspect-angle diversity is SAR (synthetic aperture radar), in which the motion of an airborne radar forms a synthetic aperture. Similarly in ISAR (inverse synthetic aperture radar), the radar is on the ground and the target is an aircraft. It is of interest that the word 'decorrelate' (p. 112) appears in S&S in the segment where they discuss the reduction of the sidelobe pattern for SAR. This suggests again that the there is a useful connection to SR.[18]

In the *frequency-diversity* method (S&S, Section 5-4), snapshots are obtained at a number of frequencies and then combined.

11.3 Single Periodic Array; the FBLP Method

In their Section 11-4, Steinberg and Subbaram describe this, the fourth, decorrelation method (of the methods listed at the start of Section 11) and apply it in the SR context. Recall that, in Section 9.1, within an ARMA model, the solution for the angular spectrum was found, subject to the need to first determine the data correlations. The discussion in S&S leads to the preferred method for estimating these correlations, namely the forward-backward linear prediction (FBLP) method.

This fourth decorrelation method may be understood physically as follows (S&S, pp. 315–317). A *periodic* array may be thought of as containing subarrays, identical with each other but *shifted* by a multiple of the inter-element spacing. *The system therefore satisfies the conditions for a single array moved to various locations, as in aspect-angle diversity* (i.e. the second of the listed methods). Only a single pulse from a single transmitter is required. This use of the periodic nature of the array to synthesise independent 'snapshots' of a scene of coherent targets is called *spatial smoothing* or *subarray smoothing*.

The array considered by S&S has spacing equal to $\lambda/2$. Recall that what remains to be done to complete the determination of the SR image in Section 9.1 is to determine the *data correlations* $R(m,n)$ defined by Equation (9.1). In synthetic decorrelation, the general method is to construct an ensemble of situations, such that the targets are uncorrelated in the ensemble. (This is quite distinct from what happens in real time.) Then an ensemble average of $e_m e_n^*$ yields $R(m,n)$. S&S describe four methods—to be called submethods—for estimating the data correlations. The first three submethods are lead-ups to the fourth, which is definitely superior. This fourth submethod is called the *modified covariance method,* or alternatively the *forward-backward linear prediction* (FBLP) method. The final result for $R(m,n)$ is given by their Equation (11-36). Armed with these estimates, one can immediately utilise the results of Section 9.1 to obtain the angular spectrum.

[17] An interesting unanswered question is how registration of the various images onto each other was achieved.

[18] Another hint of relevance to SR is that (S&S, p. 325) experimental data that S&S use to test SR methods is ISAR data.

The above theory is developed for the far field. But S&S (Section 11-6, Chapters 8 and 9) show that the near field is amenable to the same treatment, at least provided that the target is in the region where an adequate approximation is to retain just one term in the phase beyond the far-field formula. That one term is $kx^2/2r$, where $k = 2\pi/\lambda$, x is the coordinate of the target or image point measured parallel to the array, and r is the range. Then adaptive beamforming (ABF) is used to transform near-field target scenarios to the far field. Only the narrowband case is discussed.

Thus in theory the fourth submethod seems good[19]; we now turn to experimental testing. S&S (Sections 11-7 to 11-9) test the FBLP method using experimental data from an ISAR system (see Section 11.2). They compare the images produced by three methods: FBLP, MLM (i.e. Capon's method, Section 8.1) and the 'Fourier method.' The latter is essentially delay-and-add, but involves: (i) incoherent combining of images from subarrays in one comparison that S&S make, and (ii) coherent combining in another comparison. While interesting results are obtained, the data do not demonstrate any clear-cut superiority of the FBLP over the Fourier method. Indeed, the method that performs best[20] over the limited conditions considered is the coherent-combining Fourier method!

S&S (p. 334) comment on this result as follows: 'The lower limits on the model order and SNR that can make SR methods perform better than Fourier techniques on an equal data set basis have not been fully determined; it is a subject for future research.' It is found (pp. 329–330) that the model order (P in Section 9.1) must be tailored closely to the data set. 'A low model order results in loss of targets and [loss in] resolution.' However the model order must not be raised too high because: (i) 'the effective SNR of the data correlation estimates ... decreases with increasing model order for a fixed sample size,' and (ii) spurious peaks eventually appear. Further comments on this experimental test are made in Section 10.

In regard to the model order, Nash [1994, Sections 3.6, 7.1] considers how to select the model order and the assumed number of point scatterers, in the context of using the MUSIC algorithm.

11.4 Decorrelation and the Problem of a Short Averaging Time

It was noted in the context of correlation arrays (Section 4.4) that, even in the case of uncorrelated targets, the requirement for high range resolution implies that insufficient time is available in which to do the averaging (that is, the averaging of each product of signals so as to remove short-term correlations). Now, in the context of SR, since again good range resolution is required, the question comes up: Does not the same problem of insufficient time arise in SR, once the targets have been decorrelated?

---

[19] However D. Solomon [private communication] expresses the belief that that the FBLP method results in a loss of bearing resolution for at least the conventional beamformer.

[20] We pass over the question of the basis for comparison, i.e. whether one should use an 'equal aperture' or an 'equal data set' comparison (in the terminology of S&S).

To answer this question, note that, in the FBLP method, the decorrelation is *not* achieved by arranging for the phase to be shifted repeatedly in real time as the signal is continuously reflected from the target. Rather, the key to the decorrelation is that 'the phase relationships between the targets as seen by the first subarray are different from those seen by the second subarray' (S&S 1991, p. 315). Thus there is no need for any phase shift over time. Consequently it is believed that the problem raised in the previous paragraph goes away.

Let us now look again at the analogous problem for *correlation arrays* when data correlations have to be estimated. Suppose that the FBLP method (a method of synthetic averaging) can be adapted to achieve this for correlation arrays. Then again there is no shifting of the phase in real time; and so presumably again there is no problem of insufficient time.

For correlation arrays, it is difficult to draw rigorous conclusions in this area, but it seems that the *most likely* conclusions are as follows: (i) the problem of short averaging times goes away (as discussed); (ii) the number of point targets that can be accommodated in the image is quite limited and depends on the number of subarrays (shifted versions of one subarray) within each of the two 'arms' of the total array; and (iii) the signal-to-noise ratio (of the received signals before processing) must be high (since the amount of time-averaging performed is small by comparison with radio astronomy, and so this source of increasing the SNR is not available).

12. Work on General Features of Superresolution

This section briefly reviews literature making general observations on SR methods, including both comparison of methods and formulae for resolution.

Gabriel [1982] is concerned to resolve sources in cases where the sources are coherent and have unequal strengths. He discusses and applies a few SR techniques, as well as the conventional Fourier-transform technique and the 'SLC' algorithm (found to be no better than Fourier).

In the face of coherence between sources, Gabriel says that a 'Doppler-shift' method is *sometimes* available. In this method, a number of snapshots are taken while the receiving array is moved sideways. '... it is desirable to have enough snapshots to sample one or more complete Doppler cycles ... .' (The method may be the same as the use of aspect-angle diversity.) In a case considered, the improvement is 'dramatic.'

Gabriel then says: 'If a Doppler-cycle shift is not available, then we are forced to address the difficult fixed-phase coherent case.' And 'the best technique found to date' is the 'forward-backward subaperture-shift' solution. This appears to be identical to the FBLP method described by S&S. Gabriel uses simulations to test both this and an 'eigenvalue/eigenvector algorithm' derivable from the MLM (maximum likelihood method) algorithm of Capon and Lacoss (described in Childers [1978], which contains reprints of some early key papers). In the test, Gabriel compares the two methods with the Fourier-transform technique. '[The] test cases demonstrate the remarkable

improvements obtained with the recent estimation techniques, and they point to the possibilities for real-world applications.'

Clarke and Mather [1985] look at the question of *which SR technique is best.* They say that the answer varies from problem to problem, depending on the collateral knowledge. Each alternative algorithm applies a different constraint, model or assumption (see in particular their abstract). As a contribution towards the selection of an algorithm, 'we have set up a computer simulation which directly compares the performance of a selection of advanced algorithms and allows us to test basic ideas in this area quickly and easily.' Their simulation compares the following methods: MUSIC, Capon's method, maximum entropy and an algorithm due to Tufts and Kumaresan. Their graphs show that a method that is much better than another in one problem may be much worse in another.

Munier and Lacoume [1985] derive *formulae for 'practical bounds' on the angular resolving power* (angular separation of two sources that can just be distinguished). They make the interesting comment that 'all the continuous analysis methods (Capon, MEM-AR, Borgiotti-Lagunes, MUSIC) asymptotically exhibit *the same resolving power as the Pisarenko method.*' Here 'asymptotically' appears to mean 'in the limit in which a high signal-to-noise ratio is combined with a large number of snapshots (or averages).' Their overall conclusion is that 'resolving power is like the Loch Ness Monster ... . Though some people think [that they have found it], it really remains a quite elusive thing'!

13. Conclusion

The first of the two major areas discussed has been the area of *correlation arrays*—arrays based on correlation telescopes. We have investigated the cross-shaped array and other devices having a geometry appropriate to a correlation telescope. First, it has been pointed out that the cross-shaped array, when used as a *conventional* array, i.e. without a correlation operation being carried out, is quite unsatisfactory, since its beam pattern has very high sidelobes along the two principal axes. Devices consisting of two subarrays work well in radio astronomy, but only because the outputs of those subarrays—or of pairs of their component elements—are processed by correlating them.

The success of the correlation telescope suggests that in sonar one might profitably use a device based on the same principle—a correlation array. For this application to work, two main problems must be dealt with. First, one requires that the return signals from the various targets be uncorrelated. But whereas radio sources in different parts of the sky are uncorrelated, there is, unavoidably it seems, a correlation between the phases of the returns from the different sonar scatterers. The upshot is that some method for decorrelating the targets is required. In the context of superresolution discussed below, such methods exist; and it would appear that they can be adapted to correlation arrays.

The second problem is that the correlation array, at least in the form initially envisaged, requires a comfortably long time over which to average the signal products. An initial argument finds this requirement to be incompatible with the sonar's possessing the combination of good range resolution and low range sidelobes. However the argument assumes that phase shifts in real time are being utilised to remove correlations. If, on the other hand, a *synthetic* decorrelation method, such as the FBLP method, exists that can be applied to correlation arrays, it seems likely that this second problem goes away, as discussed in Section 11.4.

If these two major problems are satisfactorily dealt with, for satisfactory operation of the correlation array one must still check on two points. First, one must ensure that the required high signal-to-noise ratio is achieved. Second, in the case where discrete targets are modelled, one must ensure that the number of targets assumed is not too high, otherwise spurious targets are produced.

The second major area discussed has been *superresolution*. Several techniques for achieving SR have been briefly described, and references have been given to other techniques. The techniques briefly described include Capon's method, an autoregressive moving average (ARMA) model and a specialisation of the latter, the forward-backward linear prediction (FBLP) model.

An important pair of questions is: How good is each SR technique, and which of the several SR techniques is best for the Minefinder sonar? Unfortunately, while each method has had successes, the literature essentially fails to give general conclusions as to the resolution achieved by each method and the conditions under which it may be applied. Munier and Lacoume (see Section 12) make an attempt to answer such questions. Regarding the second question above (i.e. conditions), the first step towards answering it has been taken by Clarke and Mather [1985] (see Section 12). In particular, they say that the 'best' method varies from problem to problem, depending on the collateral or *a priori* knowledge. Each alternative algorithm applies a different constraint, model or assumption.

Undesirable features of SR—that is, conditions that must be imposed and unfortunate outcomes that can be produced—have been discussed in Section 10. Thus a high signal-to-noise ratio seems to be essential and some means must generally be found to decorrelate the returns from each pair of targets. An overzealous attempt to increase resolution often leads to distortion of the image in the ways listed in Section 10. The fixed-array-weights methods of Section 6 have further problems.

In both the area of correlation arrays and that of SR, an important requirement for the Minefinder application is the need to decorrelate the targets. Decorrelation methods, applicable in the case of SR but possibly also adaptable to correlation arrays, are discussed in Section 11. The *theory* of the FBLP method of decorrelating, discussed by S&S, looks very encouraging; on the other hand, the experimental test results presented so far are considerably less encouraging.

Finally we turn to overall prospects for the use of these two broad techniques in the Minefinder sonar context. The prospects for correlation arrays and those for superresolution are so similar that we discuss them together. First, a technique for the decorrelation of targets must be made to work in the sonar context. Most likely, one would adapt a method already developed for SR (perhaps for the radio context), such

as the FBLP method. One can be confident that a *synthetic* decorrelation method (such as FBLP or another spatial smoothing technique) disposes of a problem noted earlier: the need for a long time over which to form averages. Following success in decorrelating, one would still need to be aware of two limitations. The first is the limited number of discrete point targets that can be imaged. The second is the high signal-to-noise ratio required. At the *longest* ranges of minehunting sonar, this high SNR is not achievable, since the sonar is operating near the limit of mine detectability. The high SNR should however be achievable at closer ranges.

Much work remains to be done; but the work required should be considerably lessened if one can adapt a method already developed for *targets*, as opposed to sources, such as the methods described by S&S. It should be noted, however, that the extension of their work to the wideband context remains to be done.

This brings us to a last point: despite the vast literature pertaining to SR and correlation arrays for narrowband systems, rather little is known on these topics for wideband arrays.

14. Acknowledgements

Discussions with Rory Niland and Stuart Anstee were valuable in the preparation of this report, as were comments made by Daniel Solomon on the first draft. Remarks of Kam Lo and John Shaw were helpful. The Minefinder project was initiated by Brian Ferguson.

15. References

Blair, D.G. (2002). *Theory Pertaining to Comparison and Calibration in an Experiment to Measure Acoustic Reflection Coefficients* (AMRL Technical Note DSTO-TN-0417). Melbourne: Aeronautical and Maritime Research Laboratory

Blair, D.G. and Anstee, S.D. (2000). *Underwater Acoustic Imaging: A Simulation Program and Related Theory* (AMRL Technical Note DSTO-TN-0274). Melbourne: Aeronautical and Maritime Research Laboratory.

Brown, J.L. and Rowlands, R.O. (1959). Design of Directional Arrays. *J. Acoust. Soc. Am.*, **31**, pp. 1638–1643.

Burg, J.P. (1967). Maximum Entropy Spectral Analysis. In: *Proceedings of 37th Annual Int. Meeting, Society of Exploration Geophysicists*, Oklahoma City, Oklahoma, Oct.1967.

Burg, J.P. (1972). The Relationship between Maximum Entropy Spectra and Maximum Likelihood Spectra. *Geophys.*, **37**, pp. 375–376.

Capon, J. (1969). High Resolution Frequency-Wavenumber Spectrum Analysis. *Proc. IEEE*, **57**, pp. 1408–1418.

Charland, S.D. (1989). *An Investigation of the Extent of the Fresnel Region of Aperture Antennas with Attention to Superdirectivity* (Report no. DREO-TN-89-20). Ottawa: Defence Research Establishment.

Childers, D.G. (Ed.) (1978). *Modern Spectrum Analysis*. New York: IEEE Press.

Christiansen, W.N. and Hogbom, J.A. (1969). *Radiotelescopes*. Cambridge, U.K.: University Press.

Clarke, I.J. and Mather, J.L. (1985). Computer Simulation of Advanced Signal Processing Algorithms. In: *Array Signal Processing: an Underwater Group Conference* held at University of Technology, Loughborough, UK, 16th/17th December 1985; being: *Proceedings of the Institute of Acoustics*, **7**, Part 4; pp. 25–31.

Clarke, J. (1970). Aperture-Synthesis Technique for H.F. Ionospheric Radar. *Proc. IEE*, **117**, pp. 1633–1638.

d'Assumpcao, H.A. and Mountford, G.C. (1984). *An Overview of Signal Processing for Arrays of Receivers* (Report no. ERL-0313-EX). Adelaide: Electronics Research Laboratory.

Ferguson, B.G. (1998). Minimum Variance Distortionless Response Beamforming of Acoustic Array Data. *J. Acoust. Soc. Am.*, **104**, pp. 947–954.

Gabriel, W.F. (1982). *Tracking Closely Spaced Multiple Sources via Spectral-Estimation Techniques* (NRL Report 8603). Washington, DC: Naval Research Laboratory.

Gething, P.J.D. (1991). *Radio Direction Finding and Superresolution*, 2nd Ed. London: Peter Peregrinus.

Giuli, D. (1986). Polarization Diversity in Radars. *Proc. IEEE*, **74**(2), pp. 245–269.

Haykin, S. (Ed.) (1995). *Advances in Spectrum Analysis and Array Processing, Vol. III*. London: Prentice-Hall.

Hung, H. and Kaveh, M. (1988). Focussing Matrices for Coherent Signal-Subspace Processing. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-**36**, 1272–1281.

Kay, S.M. and Marple, S.L. (1981). Spectrum Analysis—A Modern Perspective. *Proc. IEEE*, **69**, pp. 1380–1419.

Makhoul, J. (1975). Linear Prediction: A Tutorial Review. *Proc. IEEE*, **63**, pp. 561–580.

Marple, S.L. (1987). *Digital Spectral Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall.

Munier, J. and Lacoume, J.L. (1985). Practical Bounds for the Angular Resolving-Power. In: *Array Signal Processing: an Underwater Group Conference* held at University of Technology, Loughborough, UK, $16^{th}/17^{th}$ December 1985; being: *Proceedings of the Institute of Acoustics*, **7**, Part 4; pp. 95–104.

Naidu, P.S. (1996). *Modern Spectrum Analysis of Time Series.* Boca Raton, Florida: CRC Press.

Naidu, P.S. (2001). *Sensor Array Signal Processing*. Boca Raton, Florida: CRC Press.

Nairn, D. (1968). Clipped-Digital Technique for the Sequential Processing of Sonar Signals. *J. Acoust. Soc. Am.*, **44**, pp. 1267–1277.

Nash, G. (1994). *The Application of Superresolution Techniques to Inverse Synthetic Aperture Radar Imaging*. PhD thesis, University of South Australia, Dept. of Electronic Engineering.

Nunn, D. (1985). Suboptimal Frequency Domain Adaptive Array Processing in Broadband Environments. In: *Array Signal Processing: an Underwater Group Conference* held at University of Technology, Loughborough, UK, $16^{th}/17^{th}$ December 1985; being: *Proceedings of the Institute of Acoustics*, **7**, Part 4; pp. 119–126.

Paulraj, A., Roy, R. and Kailath, T. (1986). Estimation of Signal Parameters via Rotational Invariance Techniques—ESPRIT. In: *Proceedings of 19^{th} Asilomar Conference on Circuits, Systems and Computers*, Pacific Grove, 6 Nov. 1985, pp. 83–89. New York: IEEE Press.

Perley, R.A., Schwab, F.R. and Bridle, A.H. (Eds.) (1989). *Synthesis Imaging in Radio Astronomy: A collection of lectures from the Third NRAO Synthesis Imaging Summer School.* San Francisco: Astronomical Society of the Pacific.

Pisarenko, V. (1973). The Retrieval of Harmonics from a Covariance Matrix. *Geophys. J. R. Astro. Soc.*, **33**, pp. 347–366.

Pritchard, R.L. (1953). Optimum Directivity Patterns for Linear Point Arrays. *J. Acoust. Soc. Am.*, **25**, 879–891.

Pritchard, R.L. (1954). Maximum Directivity Index of a Linear Point Array. *J. Acoust. Soc. Am.*, **26**, 1034–1039.

Procopio, L.W., Kashihara, T.K. and Czorpita, S. (1964). *Advanced Antenna Techniques.* Philco Corp Blue Bell P.

Rhodes, D.R. (1971). *On a Class of Optimum Aperture Distributions for Pattern Shaping.* Reprint, North Carolina State University, Raleigh, Dept. of Electrical Engineering.

Rissanen, J. (1989). Stochastic Complexity in Statistical Inquiry. *World Scientific*, Series in Computer Science, vol. **15**.

Schmidt, R.O. (1979). Multiple Emitter Location and Signal Parameter Estimation. *Proc. RADC Spectrum Estimation Workshop*, Griffiths AFB, Rome, New York (RADC-TR-79-63, Rome Air Development Center), pp. 243–258.

Schmidt, R.O. (1986). Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. Antennas Propagation*, **AP-34**, pp. 276–280.

Shaw, E. and Davies, D.E.N. (1964). Theoretical and Experimental Studies of the Resolution Performance of Multiplicative and Additive Arrays. *Radio & Electron. Eng.*, **28**, pp. 279–291.

Shearman, E.D.R., Bickerstaff, P. and Fotiades, L. (1973). Synthetic-Aperture Sky-Wave Radar: Techniques and First Results. In: *Proceedings of International Conference on Radar—Present and Future*, London, 23–25 October, 1973, pp. 414–421. London: Institution of Electrical Engineers.

Steinberg, B.D. (1976). *Principles of Aperture and Array System Design—Including Random and Adaptive Arrays.* New York: John Wiley.

Steinberg, B.D. and Subbaram, H.M. (1991). *Microwave Imaging Techniques.* New York: John Wiley.

Stoica, P. and Moses, R. (1997). *Introduction to Spectral Analysis.* Upper Saddle River, New Jersey: Prentice Hall.

Su, G. and Morf, M. (1983). The Signal Subspace Approach for Multiple Wide-Band Emitter Location. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-**31**, 1502–1522.

Tucker, D.G. (1961). *Space-Frequency Equivalence in Directional Arrays—with Special Reference to Superdirectivity and Reciprocity* (Report no. IEE-MG-479E). Institution of Electrical Engineers, U.K.

Urick, R.J. (1983). *Principles of Underwater Sound*, 3rd Ed. New York: McGraw-Hill.

Wang, H. and Kaveh, M. (1985). Coherent Signal-Subspace Processing for the Detection and Estimation of Angle of Arrival of Multiple Wide-Band Sources. *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-**33**, 823–831.

Appendix A: Time for Averaging

Consider the situation in radio astronomy. Even when two sources are uncorrelated, they are still correlated in the short term. Thus, when averaged over a short time, the cross terms in (4.6) (or their analogue in RA) are nonzero. For these terms to be negligible, the time for averaging must be great enough.

Our interest is in an echo-ranging sonar, requiring the signal to be either a short pulse or a coded signal such as a chirp. Then, as suggested under assumption 3c of Section 4.1, the requirement for a long averaging time may be difficult to satisfy.

The sonar situation is described, for example, by Equation (4.6). As we are studying the problem of sufficient time, let us suppose that the 'other' problem is not present, i.e. that the sonar targets are *uncorrelated*. This could be the case due to random drifts over time, either in range or in target strength; for simplicity we may assume that the drifts are in target strength. Then effectively, in (4.6), there are changes in a_b and a_c over time.

Consider first the case of a short pulse. Consider a single point target. Since the pulse interacts with the target only for a time *T*, where *T* is the duration of the pulse, *the effective time over which averaging occurs is T*. Suppose the reflections from two targets get 'out of step' in phase in a time τ. Then for decorrelation we require

$$1/f < \tau < T \tag{A.1a}$$

and probably

$$1/f << \tau << T \tag{A.1b}$$

The latter requirement forces T to be very large compared to $1/f$. Unfortunately this forces up the range resolution, which is $cT/2$. Usually in a short-pulse system, T is made equal to just a few times $1/f$; this is the case in medical imaging, for example. In such cases, Equation (A.1b) could not be satisfied. More generally, from (A.1b), decorrelation can be achieved, but at the expense of making the range resolution much larger than—perhaps *very* much larger than—the wavelength.

A coded signal such as a chirp may appear to be a way out. Consider the case where the two targets are in the same range resolution cell. For a given range resolution, *T* is now much longer than for the short pulse. Nevertheless, because the interaction is still spread over time *T*, the requirement is still (A.1a) or (A.1b). It is now easy to find values of τ satisfying (A.1b). However, this 'advance' is bought at a cost. The coded pulse at the receiving element, when cross-correlated with the transmitted signal, has its maximum value after crosscorrelation *reduced*. This is because high correlation occurs only in bursts of length τ, there being T/τ such bursts. These T/τ contributions to the sum that is the crosscorrelation, which previously added in phase, now add *incoherently*. Thus the maximum value after cross-correlation is reduced by a factor of order $(\tau/T)^{1/2}$. This reduction is unfortunate: it means that, where one would like there to be 100% correlation, the correlation is considerably less. Thus the requirement that the averaging time be sufficiently long appears to lead: (i) in the case of a short pulse, to a lack of good range resolution, and (ii) in the case of a coded pulse, to a raised level for the range sidelobes.

Appendix B: Thoughts on Application of Capon's Method to the Minefinder System

Here an attempt is made to write down a procedure by which the Capon method might be generalised to deal with the Minefinder problem. The appendix offers ideas only, and is to be regarded as quite speculative.

In the standard problem treated in Section 8.1.1, the system is passive and the sources are narrowband and far-field. We need to generalise the problem to an active system, a wideband transmitter and targets in the near field, at the same time allowing the directions to be specified by a 2-D angle (a pair of angles). Note as a preliminary that, if a coded or chirped signal is used, decoding of the received signals is to be carried out prior to other processing.

For the more general situation, we first discuss the specification of the problem—i.e. 'what parameters are to be found?' We later look at the question of how to find these parameters. At first sight, the specification of the problem appears to be: (A) Given the received voltage streams, and given any image point in the 3-D field, to determine the basic weight (real) and the delay[21] to be applied to each element. But this formulation has a worrying feature. The delay is closely tied to estimated range. If the delays can be varied by arbitrarily large amounts, a solution may be obtained in which the estimated ranges of the targets are in gross error.

While (A) may still turn out to be satisfactory, we put forward a second and a third alternative specification of the problem. The second is: (B) 'Given the received voltage streams, and given the image point in 3-D: (i) let us choose each delay to equal the time for go-and-return via the image point; (ii) despite the explicit presence of delays, let us choose the weights to be applied to each element to be *complex*; then (iii) the problem is to determine these complex weights.[22] In (ii), it is taken that the image amplitude is the sum of the *analytic* signals at the various elements, delayed as in choice (i) and weighted by the complex weight.[23]' Note that the phase factor in the complex weight is not accurately equivalent to an additional delay, since the system is wideband.

This last comment suggests the third specification (C), in which, as in (A), one determines a real weight and a delay, but with the delay now confined to some narrow interval about the delay-and-add value. The deviation Δt from the delay-and-add value could be confined to, say, $\Delta t \leq 1/2f_c$ or $\Delta t \leq 3/4f_c$, where f_c is the central frequency. But the introduction of inequalities is messy and so this third formulation (C) is not favoured.

Having specified (through (A), (B) or (C)) what parameters are to be found, we need to specify a Capon-type criterion that yields a unique solution for those parameters. Let us attempt to state what the criterion becomes when the formulation

[21] The delay is the time interval from the time of projection of the 'ping' to the time at which the voltage stream is evaluated for the purposes of beamforming.

[22] Possibly the weights should be chosen to be independent of the range of the image point, but still dependent on the 2-D angle.

[23] Note that, though the weight is complex, it is combined with a delay.

(B) is used. First, a constraint must be specified. By analogy with Section 8.1.1, the constraint imposed is that the gain equals unity at the selected image point (the point to which the array is steered). Here the 'gain' could be defined as:

$$(\text{output power from beamformer})\Big/\sum_n \int \left|e_n(t)\right|^2 dt$$

where $e_n(t)$ is the voltage stream at the nth element and the 'output power' would be taken as equal to the image intensity. As the second and final step, it remains to impose a minimisation condition. One would impose the condition that (subject to the constraint) the output power at the image point due to all targets is a minimum. If this condition leads straightforwardly to a solution via Lagrange multipliers, a generalised Capon-type method is specified. (One would still have to check that it gave satisfactory answers.)

Finally we discuss a quite different approach to applying the Capon method, namely the insertion of 'fake' targets (or 'fake' sources). Recall that the Capon method is an adaptive method, i.e. the weights depend on the signals received, and hence on the targets. R. Niland [private communication] has suggested that it would be desirable—if possible—to find a suitable set of weighting factors $\mathbf{w}(\theta)$ that do not depend on the targets; then all the weights can be precalculated, avoiding the need for a lengthy calculation in real time. It is further suggested that this goal could be achieved by putting in 'fake' targets. The signals that would be received from the collection of fake targets would be calculated and the Capon method applied to *them*, instead of to the *actual* signals.

Niland further reports that such a method has in fact been successfully applied to the passive, narrowband, azimuth-only, far-field case. In that case, when one is beamforming at azimuth angle θ, the *assumed* sources (at angles typified by θ') consist of: (i) an isotropic distribution of mutually incoherent sources (white noise, except in angle, not in time), representing interfering signals and noise; (ii) signals from angles θ' to the rear, representing turbulence in the water; and (iii) one source placed at $\theta' = \theta$, representing the source that, on occasion, will actually be present at (or very close to) θ. The adaptive augmented weights are calculated from the combined signal due to causes (i), (ii) and (iii) by the Capon method. The process is repeated at each other value[24] of θ.

While the method of fake targets or sources as described above raises some questions, the success of the passive, narrowband version of the method suggests that a similar method may work in the Minefinder context.

[24] It appears that, due to rotational invariance, the additional weighting factors are not only independent of the targets, but also remain unchanged with θ except for the effect of (ii), at least for a circular array. Here the 'additional weighting factor' is the factor in $\mathbf{w}(\theta)$ in Section 8.1.1 over and above the factor $e^{j\alpha}$ that produces steering in delay-and-add.

Appendix C: Transmitter-Location Diversity

S&S [1991] discuss this topic in Chapter 6 and Section 5-8. The theory of transmitter-location diversity differs according to whether the images from the different transmitters are to be combined coherently (complex image amplitudes added) or incoherently (image intensities added). Following S&S, we restrict attention to the case where the receiving elements and the transmitter locations all lie in the one straight line (the x axis).

For the *incoherent* case, the various transmitter locations are chosen *randomly* along the length of the array, following a uniform (rectangular) distribution. S&S investigate the conditions under which decorrelation is 'largely complete'; the latter description means that the image is close to what it would be if the targets were incoherent. It is found that the use of five transmitter locations—quite a small number—suffices to achieve this goal. S&S summarise incoherent diversity by saying that, while it leads to a small improvement in the distant sidelobe level, its primary benefits are twofold, as follows. First, the diversity intensity image (not amplitude image) is now the convolution of the actual intensity spectrum with the power pattern of the array. And second, the coherent targets are made to appear noncoherent.

With *coherent* diversity (coherent combining), from M transmitter locations and N receiving elements, an array of MN elements can be synthesised. (The transmitter locations are still random.) The amplitude beam pattern of this array is the product of the two constituent patterns. This leads to a moderate improvement in both the resolution and the near sidelobe levels, compared to the use of a single transmitter. In addition, the *distant* sidelobe levels decrease from a power value, relative to the peak, of $1/N$, to $2/MN$. Thus as long as M is at least moderately large, a tremendous improvement in the distant sidelobes is obtained. The *synthesised* array can even turn out to be not a sparse but a filled array (a filled aperiodic array).

DISTRIBUTION LIST

Suitability of Correlation Arrays and Superresolution for Minehunting Sonar

David G. Blair

AUSTRALIA

DEFENCE ORGANISATION

Task Sponsor **DGMD**

S&T Program

Chief Defence Scientist
FAS Science Policy
AS Science Corporate Management
Director General Science Policy Development
shared copy
Counsellor Defence Science, London (Doc Data Sheet)
Counsellor Defence Science, Washington (Doc Data Sheet)
Scientific Adviser to MRDC Thailand (Doc Data Sheet)
Scientific Adviser Joint
Navy Scientific Adviser
Scientific Adviser - Army (Doc Data Sheet and distribution list only)
Air Force Scientific Adviser
Director Trials

Systems Sciences Laboratory

Chief of .Maritime Operations Division (Doc Data Sheet and Distribution List Only)
A. Theobald
B.G. Ferguson
Research Leader Mine Warfare(Doc Data Sheet and Distribution List Only)
Dr Brian Ferguson, MOD Sydney: 5 copies
Gary Speechley, MOD Sydney
Stuart Anstee, MOD Sydney
John Shaw, MOD Sydney
Kam Lo, MOD Sydney
Ranjit Thuraisingham, MOD Sydney
Stuart Perry, MOD Sydney
David Liebing, MOD Edinburgh
Daniel Solomon, MOD Edinburgh
Darryl McMahon, DSTO Stirling
David Blair (Author): 5 copies

DSTO Library and Archives

Library Edinburgh 1 copy
Australian Archives
Library, Sydney

Capability Systems Staff

Director General Aerospace Development (Doc Data Sheet only)

Knowledge Staff

Director General Command, Control, Communications and Computers (DGC4) (Doc Data Sheet only)

Navy

SO (SCIENCE), COMAUSNAVSURFGRP, NSW (Doc Data Sheet and distribution list only)

Army

ABCA National Standardisation Officer, Land Warfare Development Sector, Puckapunyal (4 copies)

SO (Science), Deployable Joint Force Headquarters (DJFHQ) (L), Enoggera QLD (Doc Data Sheet only)

Intelligence Program

DGSTA Defence Intelligence Organisation

Manager, Information Centre, Defence Intelligence Organisation

Defence Libraries

Library Manager, DLS-Canberra

Library Manager, DLS - Sydney West (Doc Data Sheet Only)

UNIVERSITIES AND COLLEGES

Australian Defence Force Academy

Library

Head of Aerospace and Mechanical Engineering

Hargrave Library, Monash University (Doc Data Sheet only)

Librarian, Flinders University

OTHER ORGANISATIONS

National Library of Australia

NASA (Canberra)

AusInfo

Sonar Systems Manager, Thales Underwater Systems, 274 Victoria Rd, Rydalmere, NSW 2116 (2 copies)

OUTSIDE AUSTRALIA

INTERNATIONAL DEFENCE INFORMATION CENTRES

US Defense Technical Information Center, 2 copies

UK Defence Research Information Centre, 2 copies

Canada Defence Scientific Information Service, 1 copy

NZ Defence Information Centre, 1 copy

ABSTRACTING AND INFORMATION ORGANISATIONS

Library, Chemical Abstracts Reference Service

Engineering Societies Library, US
Materials Information, Cambridge Scientific Abstracts, US
Documents Librarian, The Center for Research Libraries, US

INFORMATION EXCHANGE AGREEMENT PARTNERS
Acquisitions Unit, Science Reference and Information Service, UK
Library - Exchange Desk, National Institute of Standards and Technology, US

OTHER ORGANISATIONS
Dr PT Gough, Department of Electrical and Electronic Engineering, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
Dr Leslie Kay (New Zealand), C/- Dr J Sendt, Thales Underwater Systems, 274 Victoria Rd, Rydalmere, NSW 2116

SPARES (5 copies)

Total number of copies: 63

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA	1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)
Suitability of Correlation Arrays and Superresolution for Minehunting Sonar	Document (U) Title (U) Abstract (U)

4. AUTHOR(S)	5. CORPORATE AUTHOR
David G. Blair	Systems Sciences Laboratory PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TN-0443	AR-012-410	Technical Note	July, 2002

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
490/6/83	NAV 97/164	DGMD	47	47

13. URL on the World Wide Web	14. RELEASE AUTHORITY
http://www.dsto.defence.gov.au/corporate/reports/DSTO-TN-0443.pdf	Chief, Maritime Operations Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DEFTEST DESCRIPTORS

Sonar arrays, Beam forming, High resolution, Radio telescopes

19. ABSTRACT

Towards the possible development of new sonars, particularly a hull-mounted minehunting sonar, work has been done in two areas. The first area concerns arrays—to be called 'correlation arrays'—that are based on the principle used by correlation telescopes in radio astronomy. The application of this principle might enable a large reduction in the number of sonar array elements. In this area, key results from the literature are reported and some concepts are elucidated with the aid of mathematics. The second area is that of superresolution (SR) and superdirectivity. The former consists of methods by which resolution is obtained that is better than that given by the Rayleigh criterion. In this second area, a more complete literature survey is carried out. A number of SR techniques are briefly described, including Capon's method and an autoregressive moving average (ARMA) model. An important problem for the application of the above techniques to sonar is that both SR and correlation arrays apparently require a high signal-to-noise ratio. For both SR and correlation arrays, the application to sonar faces a further key problem. The problem is that correlations usually exist between the return signals from different targets, and these correlations lead to artefacts and other image defects. Known techniques for 'decorrelating' targets are described.